Aalto University
School of Business

EFFECTIVE CORPORATE DATA QUALITY MANAGEMENT

Systematic Literature Review

Master's Thesis
Pekka Nurminen
26 May 2017
Information and Service Economy

Approved in the Department of Information and Service Economy

__ / __ / 20__ and awarded the grade

Author Pekka Nurminen

Title of thesis EFFECTIVE CORPORATE DATA QUALITY MANAGEMENT

Degree Master of Science in Economics and Business Administration

Degree programme Information and Service Management

Thesis advisor(s) Merja Halme; Anu Bask

Year of approval 2017 **Number of pages** 106 **Language** English

Abstract

As the entire world is in the transition becoming more and more data-driven, the quality of the data has become a major issue for individuals, organizations, governments and societies. The vast amount of data created every day has created various business opportunities, but the opportunity to use the data still varies due to the quality problems of the data. The next crucial issue in creating a more intelligent society is to standardize and develop effective corporate data quality management.

This thesis reviews previous studies on data quality management in order to study how an organization should manage its data quality. The focus is in business organizations, but the material reviewed consists of case studies from various organizations (e.g. military, government) indicating a society-wide issue.

This research conducts a systematic literature review on the existing material on data quality and data quality management. The goal of the systematic literature review is to review the material so that the review can be repeated according to an existing criteria. Originating from natural sciences, the systematic literature review is meant to reduce the personal bias of the researchers and increase the thoroughness and critical assessment. Another method used in the study is snowball linking method.

This study reviews the existing literature about managing strategic data assets. The focus points of the research are the definition and assessment of the organizational data quality, current issues in the data quality management, and data quality management. The results of the literature review are further discussed. The focus points of the discussion are the results, the possible limitations of the research and further study points.

Keywords Data Quality, Corporate Data Quality Management, CDQM

Tekijä Pekka Nurminen

Työn nimi TEHOKAS YRITYSDATAN LAADUNHALLINTA

Tutkinto Master of Science in Economics and Business Administration

Koulutusohjelma Information and Service Management

Työn ohjaaja(t) Merja Halme; Anu Bask

Hyväksymisvuosi 2017 **Sivumäärä 106** **Kieli** englanti

Tiivistelmä

Koko maailman muuttuessa yhä enemmän datan ohjaamaksi datan laatu on noussut merkittäväksi asiaksi henkilöille, organisaatioille, hallinnoille ja yhteiskunnille. Joka päivä luotu valtava datan määrä on luonut erilisia liikemahdollisuuksia, mutta data laadun ongelmat vaikuttavat suuresti mahdollisuuksiin käyttää dataa. Seuraava merkittävä asia älykkäämmän yhteiskunnan luomisessa on standardisoida ja kehittää tehokasta yrityksen datan laadunhallintaa.

Tämä Pro gradu-tutkielma tarkastelee aikaisemmin kirjoitettuja datan laadunhallinnan tutkimuksia selvittääkseen miten organisaation tulisi hallinnoida datan laatua. Tutkielma keskittyy yrityksiin, mutta tutkittu materiaali koostuu tutkimuksista, joita on tehty mitä erilaisimmille organisaatioille, kuten esimerkiksi asevoimat ja hallitukset. Tämä osoittaa että kyseessä on koko yhteiskuntaa koskettava ongelma.

Tässä tutkielmassa toteutetaan systemaattinen kirjallisuuskatsaus olemassa olevalle tutkimusmateriaalille datan laadusta ja sen hallinnasta. Systemaattisen kirjallisuuskatsauksen tarkoitus on tarkastella tutkimusmateriaalia niin että kirjallisuuskatsaus voidaan toisintaa määritettyjen kriteerien puitteessa. Systemaattinen kirjallisuuskatsaus tulee alun perin luonnontieteellisestä tutkimuksesta ja sen tarkoitus on vähentää tutkijoiden henkilökohtaisia ennakkoasenteita ja lisätä tutkimusmateriaalin kattavuutta ja kriittistä arviota. Toinen käytetty tutkimusmenetelmä on lumipallometodi.

Tämä tutkielma tarkastelee olemassa olevaa kirjallisuutta strategisen datan hallinnasta. Tutkimus keskittyy datan laadun määrittämiseen ja arviointiin, nykyisiin ongelmiin datan laadunhallinnassa ja malliin datan laadunhallintaan. Kirjallisuuskatsauksen tuloksista keskustellaan pidemmälle. Tutkielma keskittyy keskustelemaan tuloksista, tutkimuksen mahdollisista rajoitteista ja mahdollisista tulevaisuuden tutkimuskohteista.

Avainsanat datan laatu, yrityksen datan laadunhallinta, CDQM

Acknowledgements

I want to sincerely thank my thesis supervisors Merja Halme and Anu Bask who guided me through this process. Thank you so much for this learning experience that really helped me to understand scientific processes, challenge my own thinking, develop myself as a writer and learn about a fascinating topic. Also thank you for teaching me the process of a systematic literature review, a method that has taught me to be critical about the nature of scientific literature.

Contents

Acknowledgements v

1. Introduction 1

1.1 Motivation 2

1.2 Research Problem, Objectives and Questions 3

1.3 The Structure of the Thesis 4

2. Methods 5

2.1 Systematic Literature Review 5

2.2 Snowball Method 9

3. Procedure of the Study 10

3.1 Planning the Systematic Literature Review 11

3.2 Conducting the Systematic Literature Review 13

3.3 Results of the Systematic Literature Review 15

3.4 Snowball Method in Practice 20

4. Results 23

4.1 Description of the Material 23

4.2 Data Quality and its Assessment 28

4.2.1 Data, Information and Knowledge 29

4.2.2 Information Products and Roles in Information Manufacturing 30

4.2.3 Choosing a Data Quality Methodology 33

4.2.4 Total Data Quality Management Methodology 37

4.2.5 Data Quality Assessment Methodologies Results 40

4.3 Current Problems in Data Quality Management 40

4.3.1 The Literature on Data Quality Problems 41

4.3.2 Current Problems in Data Quality Management Results 43

4.4 Corporate Data Quality Management Literature Assessment 44

4.4.1 Bringing Data Quality into an Organization and Developing the Strategic Guidelines 45
4.4.2 Data Quality Management Model 48
4.4.3 Data Lifecycle Management 50
4.4.4 Metadata Management and Master Data Modeling 56
4.4.5 Data Quality Assurance 60
4.4.6 Data Integration 61
4.4.7 Cross Functions 65
4.4.8 Administration 67
4.4.9 Corporate Data Quality Management Results 70
5. Discussion 72
5.1 Corporate Data Quality Assessment 72
5.2 Common Problems in Data Quality Management 74
5.3 Managing the Corporate Data Quality 75
5.4 Summary of the Results 77
5.5 Theoretical Contributions 78
5.6 Implications to Management 81
5.7 Discussion on the Limitations, Reliability and Validity 82
6. Conclusion 85
6.1 Further Research Topics 85
References 87
Appendix 92

List of Figures

Figure 1. The structure of the systematic literature review (Tranfield et al., 2003, p. 214)

Figure 2. The structure of the systematic literature review based on Section 2.

Figure 3. The figure shows the amounts of articles published in different years.

Figure 4. The figure describes the amounts different research methods used in the chosen material.

Figure 5. The figure describes the amounts of different focus areas of the included studies.

Figure 6. Total Data Quality Management methodology in which the data is managed in a circular methodology (Wang, 1998, p. 60).

Figure 7. Master data quality management model developed by Otto et al (2012, p. 405).

Figure 8. Master Data Quality Lifecycle map (Ofner et al., 2013, p. 480).

Figure 9. ERP data quality evaluation model (Haug et al., 2009, p. 1059).

Figure 10. The key findings and topics of the systematic literature review.

List of Tables

Table 1. Searches conducted in EBSCO Business Source Complete, Emerald, JSTOR and Google Scholar

Table 2. Final articles resulting from the systematic literature review.

Table 3. Snowball-linked material

Table 4. The listed journals and proceedings used in the study

Table 5. The theoretical background of the articles.

Table 6. Articles used in section 4.2.1

Table 7. Articles used in section 4.2.2

Table 8. Comparison between manufactured products and information products (Wang, 1998, p. 59).

Table 9. Articles used in section 4.2.3

Table 10. List of data quality methodologies.

Table 11. Data quality methodology assessment table developed by Batini et al. (2009).

Table 12. Articles used in section 4.2.4

Table 13. Information quality categories and dimensions (Wang, 1998, p. 60).

Table 14. Articles used in section 4.3.1

Table 15. Articles used in section 4.4.1

Table 16. Articles used in section 4.4.2

Table 17. Articles used in section 4.4.3

Table 18. Articles used in section 4.4.4

Table 19. Requirements to be met by business metadata repository (Hüner et al., 2011, p. 369).

Table 20. Articles used in section 4.4.5

Table 21. Articles used in section 4.4.6

Table 22. Articles used in section 4.4.7

Table 23. Articles used in section 4.4.8

Table 24. “RACI chart for data quality management tasks illustrates how these five actors should be included in the process of corporate data quality management” (Wende, 2007, p. 420).

1. Introduction

Data has gained a more and more central position in our society and business world with the promise of more rational and fact-based decision-making. Falge, Otto and Österle (2012) emphasize how data has not just become a key strategic asset for singular companies, but also entire business networks and their functionality and profitability. Due to this, various researchers use the expression launched by Deloitte "Data ascends from the basement to the boardroom" (Deloitte, 2009, p. 6) which emphasizes the importance of data in the organizations. The rise of the more efficient ways to create and more intelligent ways to use data has raised the problem with the quality of data (Lucas, 2010).

Poor data quality is estimated to cost only the US enterprises alone over USD 600 billion a year (Lucas, 2010). Parssian, Sarkar and Jacob (2004) estimate that the costs of poor data quality may cost 8-12 % from the revenue of an average company. Lucas (2010) notes that despite the high cost this phenomenon is costing only inside the Unites States, the topic of data quality management is not very thoroughly studied. Ofner, Otto, and Österle (2012) study corporate data as a strategic asset that requires proper asset management strategy. Otto, Wende, Schmidt, and Osl (2007) note that companies often deal with the issue of corporate data quality by labeling it a technical issue, thus making the data asset management lack strategic business focus and therefore managing the central strategic asset in a non-optimal manner.

Often the academic research on data quality concentrates on three specific, inter-related issues: defining the quality of data through possible quality attribute, identifying issues and problems the organizations face when managing the data quality, and building models and models to better manage and control the organizational data quality (Otto et al., 2007). Wang and Strong (1996) have explored the data quality concept by studying different data quality dimensions. The studies have often been carried out by sending surveys to target groups and identifying the relevant attributes in the results. Similar studies have been made on the problems and issues of data quality inside different organizations. The current academic studies

on data quality management are often case studies and interviews on different aspects of data quality in different organizations. The research teams of University of St. Gallen have been highly productive in the studies on corporate data quality management. Otto et al. (2007) have carried out surveys, built case studies, and made research interviews on different aspects of organizational data quality management.

1.1 Motivation

The studies of Lucas (2010) and Otto et al. (2007) emphasize the lack of research and lack of deeper understanding of corporate data quality management. Lucas (2010) also notes the high costs of ignorance by noting the amount of money inefficiencies and lack of understanding cause which indicate how immediately this issue should be studied. Data must be treated as a strategic asset and organizations need an effective set of methods for managing their data assets.

There is a large volume of research material and studies on the topic of data quality management. The existing literature often uses case studies, in-depth interviews and literature reviews as research methods. Due to the large volume and case specifications of the research material, organizations may struggle in forming a plan for managing the data assets (Lucas, 2010).

When the issue of data quality is becoming more and more urgent and costly for companies around the globe, there is a need for a study that would review the existing material on models developed by various researchers for the companies to follow. This thesis aims to study the current literature on corporate data quality through means of systematic literature review. Even though various data quality studies apply literature reviews as their methods, there is a lack of a systematic literature review that would assess the existing literature in a replicable, scientific way. According to Cooper, Kibbler and Cookson (2004), the systematic literature review adds value to the existing research material by evaluating the shortcomings of the existing material "through a systematic comprehensive search strategy, data extraction, and documentation of component threats to validity" (Cooper et al., 2004, p.239). This thesis aims to use a systematic literature review as a research method in order to narrow this research gap.

Data quality has been studied by conducting various surveys and research interviews.

Understanding the current problems and issues helps organizations to identify them in their own operating environment and assess the costs of the problems. Hüner, Ofner, and Otto (2009) emphasize that corporate data quality is not solved by purchasing a piece of software, instead the issue has to be solved by restructuring the entire corporate data architecture and involving the entire organization in the process. A coherent corporate data quality management model can be constructed through defining data quality, identifying current problems and issues relating to it, and assessing the existing material on it.

1.2 Research Problem, Objectives and Questions

Lucas (2010) notes that the amount of research done about data quality attributes, concepts and tools for data quality improvement and assessment has been growing, the amount of the research on data quality management has been relatively low. The research problem of this thesis is how a company should organize its data quality management so that it supports the business functions and creates value for the company.

This thesis aims to research and review the existing literature on corporate data quality management in order to map out the ways to assess and define data quality. This thesis also aims to map out the current problems in data quality management. Finally, this thesis seeks means to manage data quality.

The definition of research problem and objectives helps to construct the research questions this thesis aims to answer. This study has three research questions:

1. How can a corporation define and assess its data quality?
2. What are the current problems in data quality management?
3. How should a corporation manage its data quality?

In order to answer the research questions, a systematic literature review is carried out. The goal of the review is to assess and review the existing material and answer the research questions. Another applied technique is the Snowball method which links further studies to the material of the systematic literature review.

1.3 The Structure of the Thesis

This study has six sections: Introduction, Research Methods, Procedure of the Study, Results, Discussion and Conclusion. Introduction introduces the research problem, objectives and questions. Section Research Methods discusses the two used research methods systematic literature review and the Snowball method. Section Procedure of the Study describes the process of material gathering. Section Results discusses the research carried out on the topic. Section Discussion links the theory in the literature review to the research questions and answers the questions. Finally, section Conclusion concludes the thesis and poses possible points for further research.

2. Methods

This section describes the methods used in the research. The first subsection describes the process of systematic literature review which is the main research method. The second subsection describes the snowball method as an additional method.

2.1 Systematic Literature Review

This section explores the process of choosing the systematic literature review as the research method and the steps of it. Based on the research questions, this thesis can be classified as qualitative research. Morgan and Smirchich (1980) explore the challenges of qualitative research. They emphasize it being an entire approach on an issue and not just a set of techniques (1980).

Bartunek, Bobko and Venkatraman (1993) provide guidelines in choosing an appropriate research method. They have studied the process of choosing a research method for the study (1993). According to them, researchers can use three requirements in choosing the method: "the need for significant methodological contributions, the need for adequate conceptual grounding, and the adherence to methodologically sound and accurate strategies" (Bartunek et al, 1993, p. 1362).

The need for significant methodological contributions requires the researcher to "demonstrate the "value-added" contribution" (Bartunek et al., 1993, p.1363). The researcher has to indicate how the chosen method provides knowledge that other possibilities do not provide. The need for adequate conceptual grounding requires the researcher who is bringing a method from another field of science to understand the conceptual grounding in another field of science. The need for adherence to methodologically sound and accurate strategies requires the researcher to accurately describe the processes of the chosen method so that the readers can understand what has been done and why (Bartunek et al., 1993).

The method chosen for this study is systematic literature review. Tranfield, Denyer, and Smart (2003) note that management research has traditionally used a narrative literature review which may not give the complete description of the relevant literature because the narrative review may lack in thoroughness. The lack in critical assessment and be influenced by the personal bias of the reviewer. Cooper et al. (2004) call for a systematic literature review

that applies a replicable, scientific and coherent process to manage the aforementioned problems. This fulfills the first requirement of the method of Bartunek et al. (1993) adding value to the research when comparing to the traditional methods.

The second requirement of Bartunek et al (1993) is to understand the scientific background of the chosen method. Both Cooper et al (2004) and Tranfield et al. (2003) emphasize that systematic literature review as a method originates from medical and natural sciences which increases the systematic and repeatable aspect to the research process. Tranfield et al. (2003) emphasize that the systematic approach improves the traditional management studies by eliminating personal bias.

This study uses systematic literature review as a research method because it fulfills the requirements of Bartunek et al. (1993). Its systematic process adds value to the traditional narrative literature review by eliminating personal bias. This is an important aspect in the data quality studies because narrative literature review is a widely used research method in the data quality research literature (see Figure 4, p. 27). The studies discovered in this research that use a narrative literature review, do not include description of the process of including the studies to the research. This allows the inclusion of the personal bias of the researchers and possible lack of thoroughness and this can be seen as a serious gap in the current data quality management literature. This study aims to use a systematic literature review in order to narrow this research gap. The aim of the review is to form a process to study the existing literature and analyze it to answer the defined research questions and eliminate personal bias and lack of thoroughness.

Lucas describes the existing literature about data quality management as “a set of concepts and roles, as well as their definitions, related to the management of data which is considered as a corporate asset” (Lucas, 2010, p. 177). She also notes that the existing literature needs clarification because both Lucas (2010) and Otto et al. (2007) have identified problems in data quality management to which the existing literature does not address like lack of business involvement and preventive strategy in the management of the data quality. This thesis aims to narrow this research gap and gather the existing material through a systematic literature review to answer the research questions about the data quality management in a clear way and address the current issues in data quality management.

Both Croom, Romano and Giannakis (2000) and Tranfield et al (2003) emphasize that a systematic literature review should have a defined structure. Tranfield et al. (2003) divide the process into three stages: planning the review, conducting a review and reporting and dissemination. The structure of the process fulfils the third requirement for adherence to methodologically sound and accurate strategies (Bartunek et al., 1993). The first stage consists of three phases: The identification of the need for a review, preparation for a proposal of review and the development of the review protocol (Tranfield et al., 2003). They emphasize the importance of a scoping study of assessing the size and the relevance of the literature in the first stage. This stage may also include a preliminary "overview of the theoretical, practical and methodological history debates surrounding the fields and sub-fields of study" (Tranfield et al., 2003, p. 214-215). The final outcome of the first stage should be a comprehensive review protocol that should specify the questions of the study, the population of the material, the search strategy, and the criteria for exclusion and inclusion of the research material.

The second phase starts with the identification of the search terms which can be based on the scoping study that has been carried out in the first phase (Tranfield et al., 2003). It is recommendable that the review would exceed further than published journals and could include conference proceedings, industry publications and unpublished studies. This phase should only include studies that comply with the established inclusion criteria. According to Tranfield et al. (2003), this phase should also include the quality assessment of the selected material. They emphasize a strict set of criteria for inclusion and exclusion that can determine the quality of the reviewed studies. The third stage of the systematic literature review process consists of reporting the outcome of the review (Tranfield et al., 2003). This procedure has been illustrated in the Figure 1.

After conducting the three stages, both Croom et al. (2000) and Gammelgaard (2004) suggest to include a section that describes the studied literature in the systematic literature review. The description can be in the form of different graphs and figures or in tables. Croom et al. (2000) note that the description is meant to give an overview of the material and to show possible patterns in it.

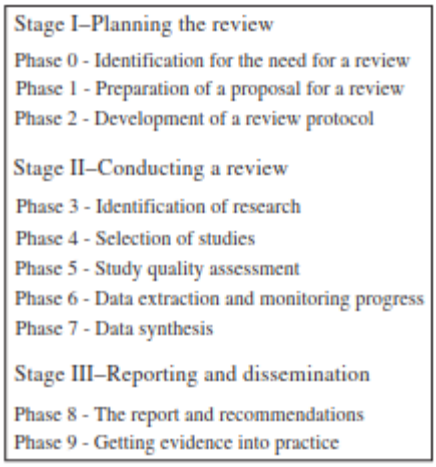

Figure 1. The structure of the systematic literature review (Tranfield et al., 2003, p. 214) divides the process of systematic literature review into three phases: planning, conducting and reporting and disseminating.

Tranfield et al. (2003) note that the method coming from a medical background, is highly suitable for quantitative research review, but qualitative research poses notable problems because it "is non-standard, unconfined, and dependent on the subjective experience of both the researcher and the researched" (Greenhalgh and Taylor, 1997, p. 741). Tranfield et al. (2003) and Greenhalgh and Taylor (1997) develop possible guideline points for defining the quality and relevance for exclusion and inclusion criteria:

- "Is the research aiming to explore the subjective meanings that people give to particular experiences and interventions" (Tranfield et al., 2003, p. 216).
- Is the research sensitive to changes that may occur during the study? (Tranfield et al., 2003)
- Can the chosen qualitative approach be considered appropriate? (Greenhalgh and Taylor, 1997)
- How did the researchers select the setting and the subjects? (Greenhalgh and Taylor, 1997)

2.2 Snowball Method

The process of gathering material can be continued with snowball method in which "one subject gives the researcher the name of another subject, who in turn provides the name of a third, and so on" (Atkinson and Flint, 2001, p. 2). Baltar and Brunet (2012) emphasize that this approach is very useful in a qualitative and descriptive research. A researcher can use this method by linking already discovered material to new material. According to Atkinson and Flint (2001), the references to the other research material can be assessed through different criteria. The list below is defined as the inclusion criteria specifically for study.

1. The material has been referred to in the material collected in the systematic literature review.
2. Relevance to the explored topic.

This study uses the snowball technique by studying the references of the studies that were included in the systematic literature review process and having the goal of finding new material relevant to the research questions. If there is an article that seems to be related to the research questions, it is searched and assessed according to the inclusion criteria above.

3. Procedure of the Study

This section describes the empirical process of conducting the systematic literature review and the snowball method. The structure of this systematic literature review has been described in Figure 2 which divides the procedure into five stages and phases in the stages. The figure combines the theories of Tranfied et al. (2003), Croom et al. (2000), Gammelgaard (2004) and Atkinson and Flint (2001). The goal of the systematic literature review is to systematically gather and analyze the material on the topic to seek answers to the defined research questions. The structure of the section is following: Subsection 3.1 describes the planning procedure of the review (Stage I), subsection 3. 2 depicts the process of conducting the review and subsection 3. 3 describes the results of the systematic literature review (Stage II). Subsection 3.4 describes the process of conducting Snowball method (Stage III). The analysis of the material can be found on Section 4 Results in subsection 4.1. (Stage IV). The reporting and recommendations have been done in Section 4 Results and in Section 5 Discussion (Stage V).

Stage I - Planning the review
- Phase 0 - Identification fore the need of the review
- Phase 1 - Preparation of a proposal fora review
- Phase 2 - Development of a review protocol

Stage II - Conducting a review
- Phase 3 - Identification of research
- Phase 4 - Selection of studies
- Phase 5 - Study quality assessment
- Phase 6 - Data extraction and monitoring process
- Phase 7 - Data synthesis

Stage III - Snowball method
- Phase 8 - Identify possible new material to be linked to the material
- Phase 9 - Search the material
- Phase 10 - Inclusion decision

Stage IV - Analysis of the included material
- Phase 11 - Analyze the gathered material
- Phase 12 - Report the analysis

Stage V - Reporting and dissemination
- Phase 13 - The report and recommendations
- Phase 14 - Getting evidence into practice

Figure 2. The structure of the systematic literature review based on Figure 1 and is adapted from it.

3.1 Planning the Systematic Literature Review

The review was started by acknowledging the need for a systematic literature review and planning the review. A scoping study was planned and conducted. The scoping study followed the idea of Tranfield et al. (2003) and provided an overview of the material to study the size and relevancy of the literature.

The scoping study was started by conducting random searches in four databases EBSCO Business Source Complete, EMERALD Insight, JSTOR and Google Scholar. The aforementioned set of four databases were chosen to get a large set of results. The reason for using four databases was to ensure that the research would not rely on only one set of search algorithms and database contents. The search term defined for the scoping study was defined to be *data quality management* because it summarizes the research questions and the topic of the study.

The first notion is that the research material is extremely large. The amount of initial search results combined from all four databases is 4,991,750 results that do overlap with each other. Tranfield et al (2003) emphasize the selection of very specified search terms.

There were three articles that were found in the scoping study phase: "Towards Corporate Data Quality Management" by Lucas (2010), "Toward a Framework for Corporate Data Quality Management" by Otto, Wende, Schmidt and Osl (2007) and "What Data Quality Means to Data Consumers" by Wang and Strong (1996). These articles were included in the material used in this study in the early stages of the research because they address the defined research questions and contain relevant material on the researched issue, data quality and data quality management.

Viewing the articles together with the research questions helped to form a set of search terms for the databases. In order to study the tools to allow a company to define its data quality, the following search terms were defined: *data quality, data quality characteristics* and *data quality attributes.* These three search terms are based on the articles of Wang and Strong (1996) and Otto et al. (2007). They can be seen favorable for studying the qualitative concept through quality characteristics and attributes.

In order to study data quality management in an organization and its current problems,

the following set of search terms were developed: *corporate data quality, corporate data quality management*, *master data* and *master data management.* These terms are based on Lucas (2010) and Otto et al. (2007) and were chosen because they specifically address the management challenges and practices in an organization.

Tranfield et al. (2003) emphasize that the systematic literature review should study not only published journal articles, but also material like conference proceedings, industry trials and unpublished trials. The databases always allow the researcher to choose the type of the publication. This study searches for all types of material from the databases to get various materials and includes conference proceedings.

The search strategy for this study is to use the aforementioned search terms with the four databases. The reason to use four databases is to avoid possible technical bias by relying on only one set of search algorithms, but instead uses four sets of search algorithms by deploying four databases. The problem of the large volume of the initial search results has to be solved because there are no resources available to study volumes that are measured in millions of hits. Because of the scarce resources, this study includes the first 500 search results of each search term in each database. The volume of 500 search hits were used because the search results were tested empirically by reviewing 100 search results after the first 500 search results in every database. The sample of 400 results after the first 500 results was highly irrelevant comparing to the search terms. The empirical testing indicated that the most relevant material was included in the first 500 results in every search and that the number of duplicates was significant across the four databases.

The four databases also sort the search results by their relevancy to the search terms. EBSCO Business Source Complete and Google Scholar provide their definitions for relevancy. EBSCO Business Source Complete searches for relevant articles by searching the exact match for the used search term and lists the results containing the exact match first. After the exact matches have been listed, they are followed by the most relevant incomplete matches that apply alphabetics in setting the results in order (EBSCO Business Source Complete). Google Scholar defines relevancy by searching for terms that are similar to the search results (Google Scholar).

Tranfield et al. (2003) note that the inclusion has several stages. This study conducts

the searches by specific search terms and decision of references, the second phase is to study the abstracts of the articles and the final stage is to read the entire article. Lucas (2010) and Wang and Strong (1996) provide guidelines for the inclusion criteria. The following set of inclusion criteria was employed:

Stage 1: Inclusion criteria for the search of the articles. The terms are searched throughout the article text and it is not limited to title or abstract.

1. The first 500 search results are included in the pool of articles.
2. The article has to be in English

Stage 2: Inclusion criteria for reading the abstracts of the articles.

The article has to be relevant to at least one of the established research questions:

1. How should a company define its data quality?
2. What are the current problems in data quality management?
3. How should a corporation manage its data quality?

Stage 3: Inclusion criteria for reading the articles.

1. The research has to explore the relevant research question and not just be related to it.
2. The chosen research method has to be reliably detailed and appropriate for the reseach objectives.

3.2 Conducting the Systematic Literature Review

The second phase of the review starts by conducting the searches with the defined search terms: *data quality, data quality characteristics, data quality attributes, corporate data quality, corporate data quality management, master data* and *master data management.* In Table 1 there are summaries of the included material during the different stages and databases.

EBSCO Business Source Complete				
Search term	Results	Stage 1 Results: First 500	Stage 2 Results: Based on the abstract	Stage 3 Results: Based on the article
data quality	10542	500	18	1
data quality attributes	34	34	1	1
data quality characteristics	61	61	1	1
data quality management	515	515	8	8
corporate data quality management	3	3	1	1
master data	774	500	5	2
master data management	355	355	2	2
total	12284	1968	36	16
Emerald				
Search term	Results	Stage 1 Results: First 500	Stage 2 Results: Based on the abstract	Stage 3 Results: Based on the article
data quality	114873	500	16	15
data quality attributes	46228	500	8	3
data quality characteristics	72189	500	9	3
data quality management	99380	500	16	2
corporate data quality management	36042	500	7	1
master data	23134	500	9	1
master data management	20511	500	7	1
total	412357	3500	72	26
JSTOR				
Search term	Results	Stage 1 Results: First 500	Stage 2 Results: Based on the abstract	Stage 3 Results: Based on the article
data quality	120032	500	19	2
data quality attributes	20573	500	5	0
data quality characteristics	56710	500	4	1
data quality management	51855	500	8	1
corporate data quality management	75777	500	10	0
master data	24586	500	1	0
master data management	9095	500	0	0
total	358628	3500	47	4
Google Scholar				
Search term	Results	Stage 1 Results: First 500	Stage 2 Results: Based on the abstract	Stage 3 Results: Based on the article
data quality	6420000	500	16	10
data quality attributes	2820000	500	8	4
data quality characteristics	5570000	500	3	3
data quality management	4840000	500	18	4
corporate data quality management	2390000	500	8	3
master data	4550000	500	7	3
master data management	3430000	500	7	3
total	30020000	3500	67	30

Table 1. Searches conducted in EBSCO Business Source Complete, Emerald, JSTOR and Google Scholar.

Stage 1 Results shows the amount limited to the first 500 search results. Stage 2 Results column shows the amount of material included on the basis of the abstracts of the studies. The Stage 3 Results shows the amount of material included on the basis of reading the material. The important aspect in the material is the large volume that caused practical difficulties. The searches conducted in four different databases resulted in 76 articles. There was overlap with the selected articles. The elimination of duplicates resulted in 31 results.

3.3 Results of the Systematic Literature Review

The three stages narrow the amount of articles into 76 articles. When the duplicates are removed, the amount of articles is 31. The articles have been listed in the Table 2.

Number	Year	Authors	Title	Journal	Approach	Focus Area
1	1993	John, S.A.	Data integration in GIS - the question of data quality	Aslib Proceedings	Empirical study: case study	Data Integration
2	1995	Wang, Richard Y.;Storey, Veda C.and Firth, Christopher P.	A Framework for Analysis of Data Quality Research Data Quality Research	IEEE Transactions on Knowledge and Data Engineering	Narrative literature review	Data Quality
3	1996	Wand, Yair and Wang, Richard Y.	Anchoring Data Quality Dimensions in Ontological Foundations	Communications of the ACM	Narrative literature review	Data Quality
4	1996	Wang, Richard Y.; Strong, Diane	Beyond Accuracy: What Data Quality Means to Data Consumers	Journal of Management Information Systems	Empirical study: survey research	Data Quality
5	1997	Hurley, Margaret A. and Harris, Rod	Facilitating corporate knowledge: building the data warehouse	Information Management & Computer Security	Narrative literature review	Data Warehouse
6	1998	Cheng, Paul S. and Chang, Pintsang	Transforming corporate information into value through data warehousing and data mining	Aslib Proceedings	Empirical study: case study	Data Warehouse
7	1998	Wang, Richard Y	A Product Perspective on Total Data Quality Management	Communications of the ACM	Empirical study: survey research	Data Quality
8	2000	Ma, Catherine; Chou, David C. ; Yen, David C.	Data warehousing, technology and assessment	Industrial Management & Data Systems	Narrative literature review	Data Warehouse
9	2002	Pipino, Leo L; Lee, Yang W.; Wang, Richard	Data Quality Assessement	Communications of the ACM	Empirical study: case study	Data Quality

Number	Year	Authors	Title	Journal	Approach	Focus Area
10	2002	Xu, Hongjiang; Nord, Jeretta Horn; Brown, Noel; Nord, G. Daryl	Data quality issues in implementing ERP	Industrial Management & Data Systems	Empirical study: case study	Database Management
11	2003	Lillrank, Paul	The Quality of Information	International Journal of Quality and Reliability Management	Narrative literature review	Data Quality
12	2004	Capiello, Cinzia; Capiello, Cinzia; Pernici, Barbara	Data quality assessment from the user's perspective	IQIS '04 Proceedings of the 2004 international workshop on Information quality in information systems	Narrative literature review	Data Quality
13	2004	Parssian, Amir; Sarkar, Sumit ; Jacob, Varghese S.	Assessing Data Quality for Information Products: Impact of Selection, Projection, and Cartesian Product	Management Science	Empirical study: case study	Data Quality
14	2007	Otto, Boris; Wende, Kristin; Schmidt, Alexander and Osl, Philipp	Towards a Framework for Corporate Data Quality Management	ACIS Proceedings	Narrative literature review	Data Quality
15	2007	Scarisbrick-Hauser, AnneMarie, and Rouse, Christina	The Whole Truth and Nothing but the Truth? The Role of Data Quality Today	Direct Marketing: An International Journal	Narrative literature review	Data Quality

Number	Year	Authors	Title	Journal	Approach	Focus Area
16	2009	Batini, Carlo; Capiello, Cinzia; Ciara Francalanci, Ciara and Maurino, Andrea	Methodologies for Data Quality Assessment and Improvement	ACM Computing Surveys	Narrative literature review	Data Quality
17	2009	Haug, Anders; Stentoft, Jan Arlbjorn; Pedersen, Anne	A Classification Model of ERP System Data Quality	Industrial Management & Data Systems	Empirical study: case study	ERP Data Quality
18	2009	Hüner, Kai M.; Martin Ofner, Martin and Otto, Boris	Towards a Maturity Model for Corporate Data Quality Management	SAC 09 Proceedings of the 2009 ACM Symposium on Applied Computing (2009)	Empirical study: case study	Corporate Data Quality Management
19	2010	Lucas, Ana	Towards Corporate Data Quality Management	Portuguese Journal of Management Studies	Empirical study: case study	Corporate Data Quality Management
20	2011	Derby, Dustin C.; Andrea Haan, Andrea and Wood, Kurt	Data quality assurance: an analysis of patient non-response	International Journal of Health Care Quality Assurance	Empirical study: survey research	Data Quality Assurance
21	2011	Hüner, Kai M.; Otto, Boris and Österle, Hubert	Collaborative Management of Business Metadata	International Journal of Information Management	Empirical study: case study	Metadata Management

Number	Year	Authors	Title	Journal	Approach	Focus Area
22	2011	Silvola, Risto; Jaaskelainen, Olli; Kropsu-Vehkapera, Hanna and Haapasalo, Harri	Managing one master data - challenges and preconditions	Industrial Management & Data Systems	Empirical study: in-depth interviews	Masterdata management
23	2012	Ofner, Martin H; Otto, Boris; Österle, Hubert	Integrating a data quality perspective into business process management	Business Process Management Journal	Empirical study: case study	Process Management
24	2012	Otto, Boris; Huner. Kai M.; Österle, Hubert	Toward a functional reference model for master data quality management	Inf. Syst E-bus Manage	Empirical study: case study	Masterdata Management
25	2013	Elbireer, Ali; Le Chasseur, Julie; Jackson, Brooks	Improving laboratory data entry quality using Six Sigma	International Journal of Health Care Quality Assurance	Empirical study: case study	Data Quality
26	2013	Haug, Anders; Albjorn, Jan Stentoft; Zachariassen, Frederik and Schlichter, Jakob	Master data quality barriers: an empirical investigation	Industrial Management & Data Systems	Empirical study: survey research	Masterdata management
27	2013	Ofner, Martin H; Straub, Kevin; Otto, Boris and Österle, Hubert	Management of the masterdata lifecycle: a framework for analysis	Journal of Enterprise Information Management	Empirical study: case study	Masterdata management

Number	Year	Authors	Title	Journal	Approach	Focus Area
28	2014	Glowalla, Paul; Sunyaev, Paul	ERP System Fit - an Explorative Task and Data Quality Perspective	Journal of Enterprise Information Management	Empirical study: in-depth interviews	ERP Data Quality
29	2014	Kwon, Ohbyung; Namyeon Lee, Namyeon and Shin, Bongsik	Data Quality Management, Data Usage Experience and Acquisition Intention of Big Data Analytics	International Journal of Information Management	Empirical study: survey research	Masterdata Management
30	2015	Huang, Hong	Domain knowledge and data quality perceptions in genome curation work	Journal of Documentation	Empirical study: survey research	Data Quality
31	2016	Aljumaili, Mustafa; Karim, Ramin and Tretten, Phillip	Metadata-based data quality assessment	VINE Journal of Information and Knowledge Systems	Empirical study: case study	Metadata, data quality

Table 2. Final articles resulting from the systematic literature review.

3.4 Snowball Method in Practice

Another chosen technique used in the literature review is snowball method, in which the references of the material found in the systematic literature review are studied to find new relevant material. When reviewing the material by employing the snowball method, the following three conditions were considered for inclusion in the current study:

1. The material has been referred to in the material collected in the systematic literature review.
2. Relevance to the explored topic.

The first inclusion criteria is the references in the material gathered in the systematic literature review. If the previously studied material referred widely to a study, it was considered potential

for inclusion. This method also helps to include the material that may have been excluded in the systematic literature review due to the rule of the first 500 search results.

When searching material on data quality and corporate data quality management, the relevance aspect was studied similarly as in the systematic literature review. The relevancy was defined by first reading the abstract of the study and if it was fit, reading the whole article. The articles found with the Snowball method have been listed in Table 3.

Elgammal, Turetken and Van Den Heuvel (2012) was chosen because it is researching regulatory compliance that was needed in the research. Compliance is one of the topics that is not covered well in the data quality management literature. This research still considers compliance controls to be an important tool in enforcing improved data quality management.

The studies of Falge, Otto and Österle (2012) and Wende (2007) were included because they belong to a group of St. Gallen University researchers that were included in the systematic literature review. The two included studies were chosen because they continue to develop the research topics and models developed by the St. Gallen studies.

The study of Miller, Malloy, Masek and Wild (2001) was chosen because it thoroughly defines *data, information* and *knowledge.* The material covered in the systematic literature review lacked a basic definition for the three basic terms so there was a need to separately search for an article that would contain the basic definition.

Year	Authors	Title	Journal	Approach	Focus Area
2012	Elgammal, Amal; Turetken, Oktay and Van Den Heuvel, Willem-Jan	Using Patterns for the Analysis and Resolution of Compliance Violations	International Journal of Cooperative Information Systems	Narrative Literature Review	Regulatorly Compliance
2012	Falge, Clarissa; Otto, Boris; Österle, Hubert	Data Quality Requirements of Collaborative Business Processes	2012 45th Hawaii International Conference on System Sciences	Case Study	CDQM
2007	Wende, Kristin	A Model for Data Governance - Organising Accountabilities for Data Quality Management	ACIS 2007 Proceedings	Narrative Literature Review	CDQM
2001	Miller, Bob; Malloy, Mary Ann; Masek, Ed and Wild, Chris	Towards a Framework for Managing the Information Environment	Information Knowledge Systems Management	Case Study	Information Management

Table 3. Snowball-linked material

The material gathered in Table 3 is combined with the material gathered in the systematic literature review. This increases the amount of included material to 35. The analysis of the 35 studies has been made in section 4.1.

4. Results

This literature review reviews the listed literature through three defined research questions:

1. How can a corporation define and assess its data quality?
2. What are the current problems in data quality management?
3. How should a corporation manage its data quality?

Section 4.1 analyzes the material gathered in the systematic literature review. Section 4.2 explores the models that allow organizations to assess and define their data quality. Section 4.3 explores the current problems in data quality management. Section 4.4 explores the literature on data quality management models.

4.1 Description of the Material

After the systematic literature review and the Snowball method the amount of included articles (Table 2 and 3 combined) is 35. This section analyses the chosen material. It analyses the years of publication, methodologies and the focus areas. The list of material included is in the Appendix.

Figure 3 describes the amounts of material published in different years. Table 4 lists the sources of the study. Figure 4 illustrates the amounts of different methods used in data quality studies. Figure 5 illustrates the amounts of different focus areas of the studies. Table 5 describes the theoretical backgrounds of the studies.

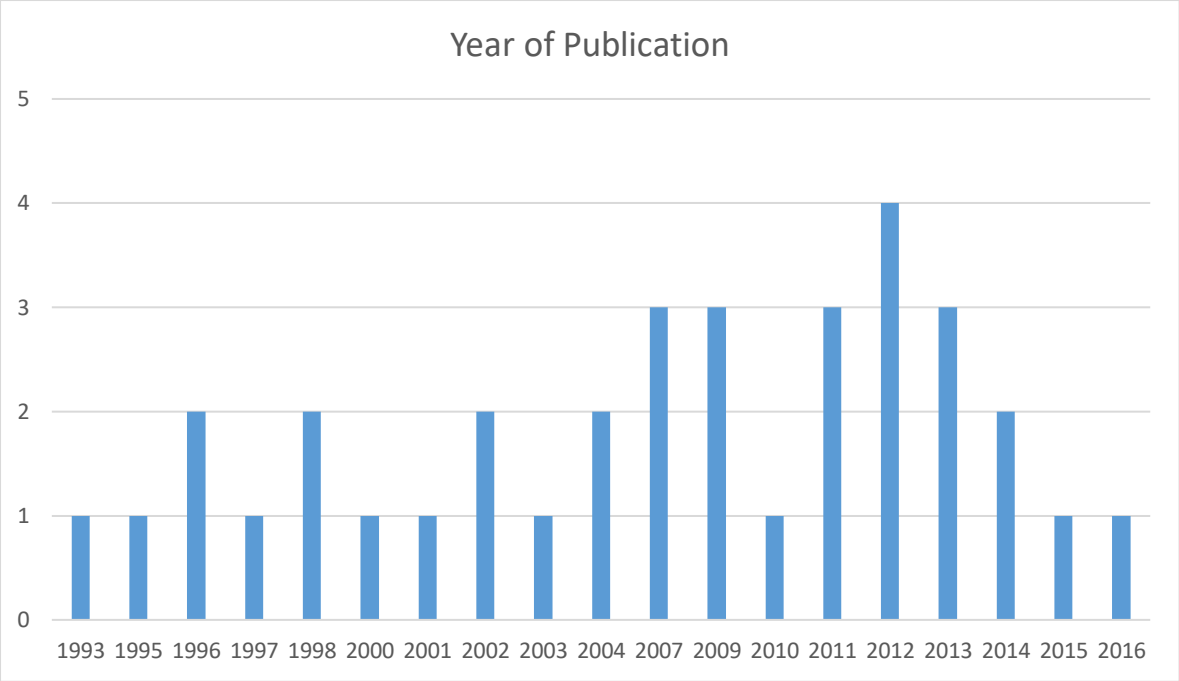

Figure 3. The figure shows the amounts of articles published in different years. Data quality has been studied throughout the computerized era of history. There has not been peak years in the data quality studies, but the amount of relevant studies published can be considered steady.

Row Labels	Count of Number
2012 45th Hawaii International Conference on System Sciences	1
ACIS 2007 Proceedings	1
ACIS Proceedings	1
ACM Computing Surveys	1
Aslib Proceedings	2
Business Process Management Journal	1
Communications of the ACM	3
Direct Marketing: An International Journal	1
IEEE Transactions on Knowledge and Data Engineering	1
Industrial Management & Data Systems	5
Inf. Syst E-bus Manage	1
Information Knowledge Systems Management	1
Information Management & Computer Security	1
Interantionl Journal of Quality and Reliability Management	1
International Journal of	1
International Journal of Health Care Quality Assurance	2
International Journal of Information Management	2
IQIS '04 Proceedings of the 2004 international workshop on Information quality in information systems	1
Journal of Documentation	1
Journal of Enterprise Information Management	2
Journal of Management Information Systems	1
Management Science	1
Portuguese Journal of Management Studies	1
SAC 09 Proceedings of the 2009 ACM Symposium on Applied Computing (2009)	1
VINE Journal of Information and Knowledge Systems	1
Grand Total	**35**

Table 4. The listed journals and proceedings used in the study.

Tranfield et al. (2003) emphasize that the systematic literature review should include not only published journal articles, but also material like conference proceedings, industry trials and unpublished trials. This study has included various types. An important notion in the material is that the material has been divided into various publications. The most common source in this research is Industrial Management & Data Systems. The sources have been gathered in table 4.

As illustrated in Table 4, this study included a various number of journals and proceedings. Some of the proceedings classify the research method and some do not. Common definitions do not exist. Because of the variance in the research methods, this study has

classified the research methods into four major categories: Empirical study: case study; Empirical Study: in-depth interview; Empirical study: survey research; and Narrative literature review. The amount of different research methods used have been illustrated in Figure 4.

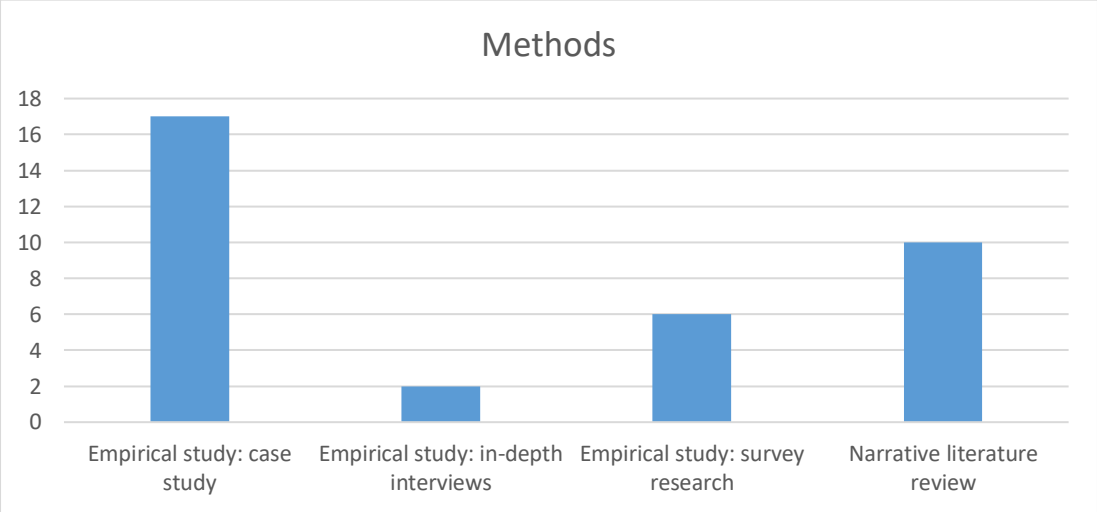

Figure 4. The figure describes the amounts different research methods used in the chosen material.

This study defines Empirical Study: Case Study a study that has been conducted for a specific company in order to answer a specific company-internal business problem. Empirical Study: in-depth interview is defined as a study in which the research method has been an in-depth interview. The difference between the two is that the in-depth interview does not include a business case, only an interview. The most used methods are literature review and case studies.

A notable feature in the literature reviews is that they have been conducted as narrative literature reviews. This indicates that the material may be affected by the personal bias of the researchers and lack of thoroughness. This further indicates that there is a research gap that could be narrowed with a systematic literature review.

The focus areas of the included material are studied in Figure 5. The categorization is based on the key words indicated in the articles and the content of the articles. The most common category in this research is Data Quality, Masterdata Management and Corporate Data Quality Management.

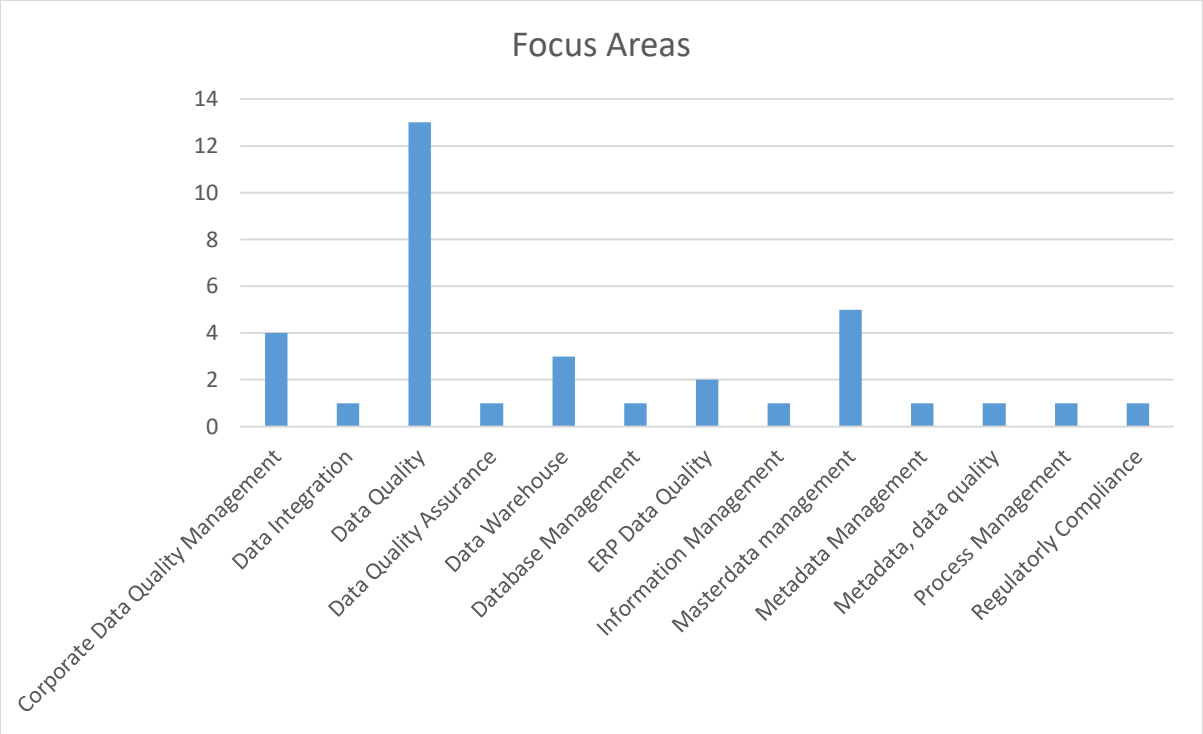

Figure 5. The figure describes the amounts of different focus areas of the included studies.

Table 5. lists the theories used in the articles. 33 of the 35 included articles were based on identifiable theories. The most used theory is the Total Data Quality Management of Wang (1998) Also, the work of Wand and Wang (1996) has been used widely. An important notion is that almost every study referred to the theory of Total Data Quality Management (Wang, 1998) which indicates the influence of the study. The list of the articles and their theories can be seen in the Appendix.

Theory	Count
A Framework for Analysis of Data Quality Research Data Quality Research (Wang, Stroey and Firth, 1995)	2
Anchoring Data Quality Dimensions in Ontological Foundations (Wand and Wang, 1996)	5
Application of Six Sigma/CAP methodology: controlling blood-product utilization and costs (Neri, Mason, Demko, 2008)	1
Beyond Accuracy: What Data Quality Means to Data Consumers (Wang and Strong, 1996)	1
Business intelligence and analytics: Form big data to big impact (Chen, Chiang, Storey, 2012)	1
Business Metadata Management (Huner, 2011)	1
Compliance	1
Data Quality in Context (Strong and Wang, 1997)	1
Improving Data Warehouse and Business Information Quality (English, 1999)	3
Mastering the information and knowledge environment (Davenport and Prusak, 1997)	1
Medical	1
Methodologies for Data Quality Assessment and Improvement (Batini, Capiello, Francallacci, Maurino, 2009)	2
N/A	2
Obtaining Information on Quality Digital Data (1986)	1
The entity-relationship model - towards a unified view of data (Chen, 1976)	1
Total Data Quality Management (Wang, 1998)	10
What Data Quality Means to Data Consumers (Wang, Strong, Guarascio, 1994)	1
Grand Total	**35**

Table 5. The theoretical background of the studies.

4.2 Data Quality and its Assessment

This subsection of the literature review gives an overview of the current academic discussion on data and data quality. It explores the research question: How can a corporation define and assess its data quality? Every subsection begins by indicating the search terms used and a table showing the articles used to explore the issue.

Subsection 4.2.1 describes the definitions of data, information and knowledge. The concepts are discussed because many researchers emphasize the relationship among the interrelated concepts. Subsection 4.2.2 describes the concept of information products and the roles relating to them. Subsection 4.2.3 explores different models of data quality assessment and a method to choose an appropriate model. Subsection 4.2.4 describes one example data

quality assessment model. Subsection 4.2.5 concludes the data quality assessment section.

4.2.1 Data, Information and Knowledge

In order to study data quality management, the terms *data*, *information*, and *knowledge* have to be defined because they are explored various times in the literature. This subsection presents two views on the terms. One explores the terms as a result of a process and the other one from the perspective of the users. Search terms used in this subsection are *data quality, data quality characteristics* and *data quality attributes.*

Articles used in this subsection are listed in Table 6.

Year	Authors	Title	Journal	Approach	Focus Area
1998	Wang, Richard Y	A Product Perspective on Total Data Quality Management	Communications of the ACM	Empirical study: survey research	Data Quality
2001	Miller, Bob; Malloy, Mary Ann; Masek, Ed and Wild, Chris	Towards a Framework for Managing the Information Environment	Information Knowledge Systems Management	Empirical study: case study	Information Management
2004	Parssian, Amir; Sarkar, Sumit ; Jacob, Varghese S.	Assessing Data Quality for Information Products: Impact of Selection, Projection, and Cartesian Product	Management Science	Empirical study: case study	Data Quality
2010	Lucas, Ana	Towards Corporate Data Quality Management	Portuguese Journal of Management Studies	Empirical study: case study	Corporate Data Quality Management

Table 6. Articles used in subsection 4.2.1

Miller, Malloy, Masek and Wild (2001) derive their definition from military systems studies. They explore the three terms from a process view in which the three terms, *data*, *information*, and *knowledge,* are considered as results of a process. In this process *information* is created from *data* and *knowledge* is created from *information.*

They define *data* as representation of objects that is based on observed facts on the researched environment. This representation may be in various forms like text, number series, sound recordings, images and film recordings. An important element of data is that it is useless for the user before it has been interpreted by a certain process that turns it into *information.*

(Miller et al., 2001)

Miller et al. (2001) continue defining *information* as the aggregation of *data* that has been given a meaning through a process that interprets that *data*. Unlike data, *information* is always meant to contain a message that is meant to affect the receiver of the message (Miller et al., 2001). This message may also be in different forms like text, number series, sound recordings, images and film recordings (Miller et al., 2001).

Knowledge can be defined as information that has been associated with a context, experience or certain rules. The information can be turned into *knowledge* through a learning process in which the association has been carried out. The form that *knowledge* often assumes differs from *data* and *information*. *Knowledge* is often considered to be a result of a learning process, so it often takes a form of an organizational routine, practice or rule. (Miller et al., 2001)

Lucas (2010) on the contrary explores *data*, *information* and *knowledge* from the user point of view. According to her *data* is a set of stored representation of objects that have a meaning for the user whereas *information* is a set of processed data that is meant to increase the user's knowledge about the object. *Knowledge* is considered a set of *information* that can be used to make decisions about the object.

These two perspectives are important in the data quality literature because the literature often explores the concepts from the process and user perspectives. Various researchers explore *data* and its quality as a result of a process and Parssian, Amir and Jacob (2004) note that if the data has low quality, the *information* and *knowledge* are going to have quality issues. On the other hand, Wang (1998) emphasizes that different users may have different needs for *data*. These concepts are explored further in the following section.

4.2.2 Information Products and Roles in Information Manufacturing

This subsection explores the data manufacturing processes and how they affect the data. The section also discusses about the different tasks related to the processes. Search terms used in this subsection are *data quality, data quality characteristics* and *data quality attributes.*

Articles used in this subsection are listed in Table 7.

Year	Authors	Title	Journal	Approach	Focus Area
1995	Wang, Richard Y.;Storey, Veda C.and Firth, Christopher P.	A Framework for Analysis of Data Quality Research	IEEE Transactions on Knowledge and Data Engineering	Literature review	Data Quality
1996	Wand, Yair and Wang, Richard Y.	Anchoring Data Quality Dimensions in Ontological Foundations	Communications of the ACM	Literature review	Data Quality
1996	Wang, Richard Y. and Strong, Diane	Beyond Accuracy: What Data Quality Means to Data Consumers	Journal of Management Information Systems	Empirical study: survey research	Data Quality
1998	Wang, Richard Y	A Product Perspective on Total Data Quality Management	Communications of the ACM	Empirical study: survey research	Data Quality
2004	Parssian, Amir; Sarkar, Sumit ; Jacob, Varghese S.	Assessing Data Quality for Information Products: Impact of Selection, Projection, and Cartesian Product	Management Science	Empirical study: case study	Data Quality

Table 7. Articles used in subsection 4.2.2

Wang, Storey and Firth (1995) note that various data quality studies study the topic through a manufacturing aspect. In these studies the creation of data is seen as a process that produces data that is called information products. Wand and Wang (1996) note that the information products' quality is depends on the data manufacturing processes that create the information products. The quality of the information products is also affected by the capabilities and needs of different people who are related to the process. Wang (1998) has described the roles of the people related to the process.

This subsection explores information products (IP) and different roles that are related to them: information suppliers, information manufacturers, information consumers, and information product managers. The listed concepts are often discussed in the data quality literature because they are the roles that create and consume the data. The different authors have used terms "data product" and "information product" interchangeably, and the current study chooses to use the term information product due in accordance with Wang (1998). This is done because Wang's studies are some of the most referred data quality studies in the field.

Parssian et al. (2004) consider an information product to be a product that has been produced in an information manufacturing process that includes different information systems. This is also supported by Wang (1998) who compares information product manufacturing to

product manufacturing. The similarities and dissimilarities between the two processes that are gathered in Table 8. According to Wang (1998), the processes of manufacturing physical products and information products have similarities: Both processes have inputs, they have processes and process outputs. In the case of a physical product, the input is a set of raw materials, there is a process for the raw materials and there is a physical product as an output. When an information product is considered, the input can be seen to be a set of raw data, there is a process for the raw data and there is an information product as an output.

Wang (1998) also explores the important differences between physical products and information products. According to him, there is a significant difference in the inputs of the process. "Data can be utilized by multiple consumers and not depleted, whereas raw material can only be used for a single physical product" (Wang, 1998, p. 58).

Products vs. information manufacturing

	Product Manufacturing	Information Manufacturing
Input	Raw Materials	Raw Data
Process	Assembly Line	Information System
Output	Physical Products	Information Products

Table 8. Comparison between manufactured products and information products (Wang, 1998, p. 59).

Wang (1998) discusses four roles in information manufacturing: information suppliers, information manufacturers, information consumers, and information product managers. "Information suppliers are those who create or collect data for the IP" (Wang, 1998, p. 60). Wang continues by defining information manufacturers as "those who design, develop, or maintain the data and systems infrastructure for the IP" (Wang, 1998, p. 60). Further "information consumers are thosa who use the IP in their work" (Wang, 1998, p. 60). Wang and Strong (1996) explore the function of information consumers with the notion "most data consumers are not purchasing data, they are choosing to use or not use data in a variety of

tasks" (Wang and Strong, 1996, p. 8). "IP managers are the ones responsible for managing the entire IP production process throughout the IP life cycle" (Wang, 1998, p. 60).

4.2.3 Choosing a Data Quality Methodology

This subsection explores the listed literature about data quality assessment. It first discusses the definition of data quality and then lists the most widely referred data quality assessment methodologies. Finally a tool to compare the listed methodologies is presented. Search terms used in this subsection are *data quality* and *corporate data quality management.*

Articles used in this subsection are listed in Table 9.

Year	Authors	Title	Journal	Approach	Focus Area
2002	Pipino, Leo L; Lee, Yang W.; Wang, Richard	Data Quality Assessement	Comunications of the ACM	Narrative Literature review	Data Quality
2003	Lillrank, Paul	The Quality of Information	Interantionl Journal of Quality and Reliability Management	Narrative Literature review	Data Quality
2004	Capiello, Cinzia; Capiello, Cinzia; Pernici, Barbara	Data quality assessment from the user's perspective	IQIS '04 Proceedings of the 2004 international workshop on Information quality in information systems	Narrative Literature review	Data Quality
2009	Batini, Carlo; Capiello, Cinzia; Ciara Francalanci, Ciara and Maurino, Andrea	Methodologies for Data Quality Assessment and Improvement	ACM Computing Surveys	Narrative Literature review	Data Quality
2012	Ofner, Martin H; Otto, Boris; Österle, Hubert	Integrating a data quality perspective into business process management	Business Process Management Journal	Empirical study: case study	Data Quality, Business Process Management
2012	Otto, Boris; Huner. Kai M.; Österle, Hubert	Toward a functional reference model for master data quality management	Inf. Syst E-bus Manage	Empirical study: case study	Data Quality Management, Masterdata Management

Table 9. Articles used in subsection 4.2.3

There is a large volume of material written on data quality and related data quality management methodologies. Otto, Hüner, and Österle (2012) provide a simple definition for overall data quality: "Despite all differences, there is consensus in the specification of the term data quality as the data's fitness for use" (Otto et al., 2012, p. 398). Various other researchers use the term

fitness for use and often indicate that it is a commonly accepted general definition for data quality. Lillrank (2003) Pipino, Lee and Wang (2002) emphasize the contextuality of data quality and note that no general "one size fits all" solution exists for defining and measuring data quality.

Both Capiello, Francalanci and Pernici (2004) and Pipino et al. (2002) point out that the current literature considers data quality to be a highly multi-dimensional concept. Both papers study the literature on data quality and have concluded that the qualitative concept of data quality is often explored through different data quality dimensions. Capiello et al. (2004) list six influencial methodologies of data quality dimensions: accuracy, completeness, consistency, timeliness, interpretability and accessibility.

Capiello et al. (2004) emphasize that the current literature on data quality does not include a comprehensive set of tools for corporations to assess and define their data quality stance. They do point out that one of the most problematic topics with data quality characteristics is the concentration on data values and not concentrating on the usage of the data. Also the process of producing the data is not often studied. This is also strongly emphasized by Ofner, Otto and Österle (2012).

Batini, Cappiello, Francalaci and Maurino (2009) note that there are various data quality methodologies produced by different researchers. These methodologies are defined as set of guidelines and techniques the purpose of which is to outline "a rational process to assess and improve the quality of data" (Batini et al., 2009, p. 2). They note that a data quality methodologies often consist of three basic phases: *State reconstruction* in which the collecting of preliminary data on processes and systems is collected, *Assessment / Measurement* that measures "the quality of data collections along relevant quality dimensions" (Batini et al., 2009, p. 2) and *Improvement* in which the procedures to achieve new data quality targets are chosen. The most common methodologies have been listed in Table 10.

Methodology Acronym	Extended Name	Main Reference
TDQM	Total Data Quality Management	Wang 1998
DWQ	The Datawarehouse Quality Methodology	Jeusfeld et al. 1998
TIQM	Total Information Quality Management	English 1999
AIMQ	A methodology for information quality assessment	Lee et al. 2002
CIHI	Canadian Institute for Health Information methodology	Long and Seko 2005
DQA	Data Quality Assessment	Pipino et al. 2002
IQM	Information Quality Measurement	Eppler and Münzenmaier 2002
ISTAT	ISTAT methodology	Falorsi et al 2003
AMEQ	Activity-based Measuring and Evaluating of product information Quality (AMEQ) methodology	Su and Jin 2004
COLDQ	Loshin Methodology (Cost-effect Of Low Data Quality	Loshin 2004
DaQuinCIS	Data Quality in Cooperative Information Systems	Scannapieco et al. 2004
QAFD	Methodology for the Quality Assessment of Financial Data	De Amicis and Batini 2004
CDQ	Comprehensive methodology for Data Quality management	Batini and Scannapieco 2006

Table 10. List of data quality methodologies. The lefternmost column shows the acronym of the methodology, the second column shows the exact name of the methodology, and the third column shows the researchers (Batini et al., 2009, p. 12).

Considering that there is no ultimate one solution for organizational data quality management issues, Batini et al (2009) have developed a method that allows organizations to assess different methodologies and determine the one that would suit for their specific situations. They define the basic steps for data quality assessment: *data analysis, data quality requirements analysis, identification of critical areas, process modelling* and *measurement of quality*. They have constructed a table (Table 11) that shows which methodologies have included those steps. This table can be considered as a tool to assess the data quality assessment methodologies and find the most optimal one.

Table 11 presents the data quality methodologies presented in Table 10 in the left colum, Step/ Meth Acronym. The full methodology names can be seen in Table 10 in the same order as in Table 11. *Data analysis* is defined as a process that "examines data schemas and performs interviews to reach a complete understanding of data and related architectural and management rules" (Batini et al., 2009, p. 4). According to them, *data quality requirements analysis* concentrates on monitoring the points the users of data and its administrators and spot possible quality issues and set new quality goals. *Identification of critical areas* chooses the

most relevant databases and data flows and quantitatively assesses them" (Batini et al., 2009).

Process modeling "provides a model of the processes producing or updating the data" (Batini et al., 2009, p. 4). *Measurement of quality* "selects the quality dimensions affected by the quality issues identified in the data quality requirements analysis step and defines corresponding metrics" (Batini et al., 2009, p. 4). Batini et al. (2009) note that the measurement may be objective and rely on quantitative metrics, or subjective and rely on qualitative evaluations by the administrators and users of the data. The last column *Extensible to Other Dimension and Metrics* indicates whether the dimensions of the methodologies can be extended to other dimensions (open) or whether the dimensions have strict definitions and cannot be extended to other dimensions (fixed).

Step/Meth Acronym	Data Analysis	DQ Requirement Analysis	Identification of Critical Areas	Process Modeling	Measurement of Quality	Extensible to Other Dimensions and Metrics
TDQM	+		+	+	+	Fixed
DWQ	+	+	+		+	Open
TIQM	+	+	+	+	+	Fixed
AIMQ	+		+		+	Fixed
CIHI	+		+			Fixed
DQA	+		+		+	Open
IQM	+				+	Open
ISTAT	+				+	Fixed
AMEQ	+		+	+	+	Open
COLDQ	+	+	+	+	+	Fixed
DaQuinCIS	+		+	+	+	Open
QAFD	+	+	+		+	Fixed
CDQ	+	+	+	+	+	Open

Table 11. Data quality methodology assessment table developed by Batini et al. (2009). This table indicates which data quality assessment steps are included by the most prominent data quality assessment methodologies (Batini et al., 2009, p. 12).

4.2.4 Total Data Quality Management Methodology

Total Data Quality Management (TDQM) developed by Wang (1998) was chosen to be an example of a data quality management methodology. Wang's work was chosen due to the high number of references to his work. His methodology for data quality management of information products is circular. Wang and Strong (1996) defined data quality dimensions that are also used in the methodology for Total Data Quality Management. Another reason Wang's methodology was chosen is that TDQM includes most steps in the assessment tool of Batini et al. (2009). Search term used in this subsection is *data quality*.

Articles used in this subsection are listed in Table 12.

Year	Authors	Title	Journal	Approach	Focus Area
1996	Wang, Richard Y. and Strong, Diane	Beyond Accuracy: What Data Quality Means to Data Consumers	Journal of Management Information Systems	Empirical study: survey research	Data Quality
1998	Wang, Richard Y	A Product Perspective on Total Data Quality Management	Communications of the ACM	Empirical study: survey research	Data Quality
2009	Batini, Carlo; Capiello, Cinzia; Ciara Francalanci, Ciara and Maurino, Andrea	Methodologies for Data Quality Assessment and Improvement	ACM Computing Surveys	Literature review	Data Quality

Table 12. Articles used in subsection 4.2.4

Wang and Strong (1996) note that data quality studies and practical implementations have been concentrating on data accuracy. They aimed "to develop a framework that captures the aspects of data quality that are important to data consumers" (Wang and Strong, 1996, p. 5). Their study "Beyond Accuracy: What Data Quality Means to Data Consumers" studies data quality through its attributes that are attributes that the data consumers thought when they think about data quality.

To reach the goal they sent a survey to various professionals dealing with data quality. The responses allowed them to form a set of categories of data quality dimensions that were considered to be important to data consumers. The four categories identified are intrinsic data quality, accessibility data quality, contextual data quality, and representational data quality. (Wang and Strong, 1996)

Intrinsic data quality category consists of the dimensions of accuracy, objectivity, believability and reputation. Wang and Strong (1996) emphasize that data consumers consider believability and reputation to be an essential part of intrinsic DQ. Accessibility data quality category contains dimensions of access and security. They have described this to be the common focus point of data quality studies. "However, there is little difference between treating accessibility DQ as a category of overall data quality, or separating it from other categories of data quality" (Wang and Strong, 1996, p. 21).

The contextual data quality category contains the following dimensions: relevancy, value-added, timeliness, completeness, and the amount of data. "Grouping of dimensions for contextual DQ revealed that data quality must be considered within the context of the task at hand" (Wang and Strong, 1996, p. 20).

Representational data quality category is formed of aspects that are linked with the the data format and the meaning of data. The data quality dimension of this category are interpretability, ease of understanding, concise representation, and consistent representation (Wang and Strong, 1996). The categories and their related dimensions are gathered in Table 13.

IQ Category	IQ Dimensions
Intrinsic IQ	Accuracy, Objectivity, Believability, Reputation
Accessibility IQ	Access, Security
Contextual IQ	Relevancy, Value-Added, Timeliness, Completeness, Amount of data
Representational IQ	Interpretability, Ease of understanding, Concise representation, Consistent representation

Table 13. Information quality categories and dimensions (Wang, 1998, p. 60).

Wang (1998) develops a TDQM cycle to be the process for data quality management of information products. The process consists of the listed phases: define, measure, analyze and improve. The cube in the middle of the circle describes the three dimensions of the managed information product: information product characteristics, information product quality and information product. The circular methodology has been described in figure 6 below.

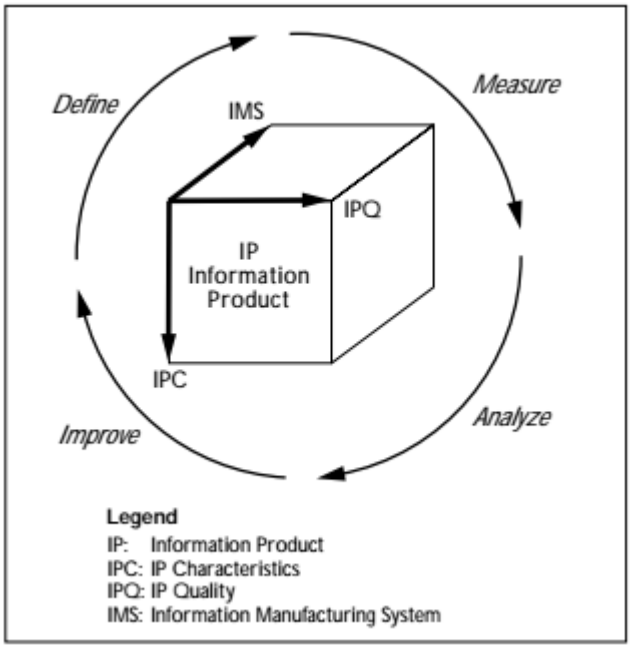

Figure 6. Total Data Quality Management methodology in which the data is managed in a circular methodology (Wang, 1998, p. 60).

TDQM method starts by defining the information product. This is done by defining the IP characteristics, information quality requirements and information manufacturing system. The IP characteristics are determined by defining the functionalities of the IP for information consumers. Information quality requirements are determined "from the perspective of the information producers" (Wang, 1998, p. 62). Wang (1998) suggests using analytics and survey tools to carry out surveys for information producers. Information manufacturing system needs be defined to gain better understanding of the system in which the data is created. If this is compared to Table 11 Data quality assessment table, this can be seen as the data analysis step.

Wang (1998) continues that the following phase in TDQM cycle is measurement. The quality of the information product needs to be somehow measured to develop it. The data quality dimensions developed by Wang and Strong (1996) can be used in this phase in order to define the measurement scale. When compared to Table 11, this can be seen as measurement of quality step.

The next phase in the cycle is the analysis of the information product. The goal of this phase is to find the root causes for possible quality problems. There are different ways the

analysis can be done. Dummy accounts could be used in order to spot the sources of the problems. It is also emphasized that the analysis has to have a targeted payoff. When compared to Table 11, this can be seen as process modelling. (Wang, 1998)

The goal of the last improvement phase is "to identify key areas of improvement" (Wang, 1998, p. 65). This can be done by "aligning information flow and work flow with the corresponding information manufacturing system" (Wang, 1998, p. 65) and "realigning the key characteristics the IP business needs" (Wang, 1998, p. 65). When compared to Table 11, this can be seen as identification of critical areas.

4.2.5 Data Quality Assessment Methodologies Results

This subsection sought to answer the research question: How can a company define and assess its data quality? The answer to this research question was sought through the means of systematic literature review that contrary to narrative literature review, applies a systematic and repeatable process to gather the material thus eliminating possible personal bias and lack of thoroughness. This approach adds value to the study because the studies included in this research always apply narrative literature review. The systematic approach narrows the existing research gap caused by the narrative literature reviews.

The research was conducted by first defining key concepts that are used in the literature often. The concepts of *data, information* and *knowledge* were defined first to study the link between the concepts. After this the information manufacturing process was presented in order to explore the process that creates the data. The definition of data quality was explored after that.

Batini et al. (2009) develop a set of assessment criteria that allow an organization assess different methodologies for the specific needs of the organization. The assessment is summarized in Table 11 and the table can be used in assessing the different data quality methodologies. After this, one example methodology was presented.

4.3 Current Problems in Data Quality Management

This section explores the research question: What are the current problems in data quality management. The subsection begins by indicating the search terms used and a table showing

the articles used to explore the issue. Subsection 4.3.1 explores the material written on the topic. Subsection 4.3.2 concludes the section.

4.3.1 The Literature on Data Quality Problems

This subsection reviews the existing literature about the current problems of the data quality management. The section first discusses the problems in high level and then presents a list of the implications of the major problems. Search terms used in this subsection are *data quality, corporate data quality, corporate data quality management* and *master data.*

Articles used in this subsection are listed in Table 14.

Year	Authors	Title	Journal	Approach	Focus Area
2003	Lillrank, Paul	The Quality of Information	Interantionl Journal of Quality and Reliability Management	Narrative Literature review	Data Quality
2007	Otto, Boris; Wende, Kristin; Schmidt, Alexander and Osl, Philipp	Towards a Framework for Corporate Data Quality Management	ACIS Proceedings	Narrative Literature review	Data Quality
2010	Lucas, Ana	Towards Corporate Data Quality Management	Portuguese Journal of Management Studies	Empirical study: case study	Corporate Data Quality Management
2011	Hüner, Kai M.; Otto, Boris and Österle, Hubert	Collaborative Management of Business Metadata	International Journal of Information Management	Empirical study: case study	Data Quality Management, Metadata Management
2011	Silvola, Risto; Jaaskelainen, Olli; Kropsu-Vehkapera, Hanna and Haapasalo, Harri	Managing one master data - challenges and preconditions	Industrial Management & Data Systems	Empirical study: in-depth interviews	Masterdata management
2013	Haug, Anders; Albjorn, Jan Stentoft; Zachariassen, Frederik and Schlichter, Jakob	Master data quality barriers: an empirical investigation	Industrial Management & Data Systems	Empirical study: survey research	Data quality, master data management

Table 14. Articles used in subsection 4.3.1

Lucas (2010) notes that even though data quality has been researched for a long time, the corporate data quality and possible corporate data quality management models have not been studied comprehensively. The scientific discussion raises two main topics on corporate data

quality management: the shift from reactive to preventive activities and the organizational change of involving the entire organization in the management of data quality. Various academics have noted that the activities have been concentrated for too long on the reactive activities. People responsible for organizational data quality have been merely concentrating on periodically fixing the quality of existing data in an existing system (Hüner et al., 2011; Silvola, Jääskeläinen, Kropsu-Vehkaperä and Haapasalo, 2011). Lucas (2010) and Hüner et al. (2011) recommend a preventive measures for corporate data quality management that can help to manage data throughout its entire lifecycle and would thus create a system that would produce quality data.

Silvola et al. (2011) note that many companies shift between reactive and passive approaches in different stages of different IT projects. This means that when a problem emerges, it is fixed with a reactive activity and then the companies often turn to do nothing on the data quality. They call for better data quality surveillance to prevent the problems.

Another major problem in corporate data quality management is the lack of business involvement in the process of managing the data quality. Lillrank (2003) emphasizes that the data requires a context in order to have some level of quality. Lucas (2010) notes that the data quality issue has been traditionally seen as an IT issue, which is problematic because the data is meant to be used by the business. Studies fail to combine the technical- and business-related aspects into one single, integrated model.

Otto et al. (2007) emphasize that there is an "interrelationship between IT and business goals of a company" (Otto et al. 2007, p. 919). Noting this misrepresentation Otto et al. (2007) form a data quality management model that aligns IT with business strategy and includes three layers: strategy, organization and information systems. These layers are studied through two organizational perspectives: governance and execution.

When the corporate data quality management is the responsibility of the technical IT employees who merely concentrate on reactive data quality management activities, the result is often the lack of strategy which leads to a situation in which the full potential of the data is not used. Otto et al. (2007, p. 917) state "neither relevant design objects are identified nor organisational responsibilities defined which is of particular importance when addressing

DQM in a cross-divisional and corporate context." Another problem is the lack of strategic orientation in the existing concepts that would include the business view of data quality management. Also, "methods and tools for DQM realisation as well as operative measures for data quality are not sufficiently specified" (Otto et al., 2007, p.917). This issue with the business view is still elaborated by noting that "existing [theoretical] concepts of DQM fall short of incorporating a long-term, strategic orientation, identifying relevant design objects, and defining organisational responsibilities" (Otto et al., 2007, p. 919).

Haug, Arlbjorn, Zachariassen and Schlichter (2013) have studied the barriers of corporate data quality management by conducting surveys on current challenges occurring in organizational data quality management. The study can be considered to be a summary of other current data quality problems. Haug et al. (2013, p. 246) have classified the following twelve data quality barriers:

1. "Missing placement of responsibilities for specific types of master data."
2. "Lack of clarity in roles in relation to creation, use, and maintenance."
3. "Inefficient organizational procedures."
4. "Lack of management focus in relation to data quality."
5. "Lack of data quality measurements."
6. "Lack of rewards/reprimands in relation to data quality."
7. "Lack of training and education of data users."
8. "Lack of written data quality policies and procedures."
9. "Lack of emphasis on the importance of data quality from managers."
10. "Lack of IT systems for data management."
11. "Lack of possibilities for input in existing IT systems."
12. "Poor usability of IT systems."

4.3.2 Current Problems in Data Quality Management Results

This section sought to answer the research question: What are the current problems in data quality management. The answer to this research question was sought through the means of systematic literature review that contrary to narrative literature review, applies a systematic and repeatable process to gather the material thus eliminating possible personal bias and lack

of thoroughness. This approach adds value to the study because the studies included in this research always apply narrative literature review. The systematic approach narrows the existing research gap caused by the narrative literature reviews.

After a careful review of the material, two main problems can be identified: the lack of preventive management and measures and the lack of business involvement in the data quality management. The shortcomings in organizational data quality management, listed in the previous subsection, can be seen as symptoms of the two main issues.

4.4 Corporate Data Quality Management Literature Assessment

This section assesses the existing literature on corporate data quality management and answers the research question: How should a corporation manage its data quality? Every subsection begins by indicating the search terms used and a table showing the articles used to explore the issue.

Subsection 4.4.1 defines important terms and concepts related to the issue. Subsection 4.4.2 gives an overview of the model and sections 4.4.3 - 4.4.8 assess the different parts of the model and add further theory into it. Subsection 4.4.9 concludes the section.

4.4.1 Bringing Data Quality into an Organization and Developing the Strategic Guidelines

This section defines several important concepts: *corporate data philosophy, corporate data policy, corporate data strategy, data governance, data quality management, data owner, data quality methodology, data quality techniques, data quality tools, master data, master data quality management* and *data steward*. According to Lucas (2010), the terms are vital for understanding and constructing a coherent long-term plan and goals for an organizational data usage. Search terms used in this subsection are *data quality, corporate data quality, corporate data quality management, master data* and *master data management.*

Articles used in this subsection are listed in Table 15.

Year	Authors	Title	Journal	Approach	Focus Area
1998	Wang, Richard Y	A Product Perspective on Total Data Quality Management	Communications of the ACM	Empirical study: survey research	Data Quality
2007	Otto, Boris; Wende, Kristin; Schmidt, Alexander and Osl, Philipp	Towards a Framework for Corporate Data Quality Management	ACIS Proceedings	Literature review	Data Quality
2007	Wende, Kristin	A Model for Data Governance - Organising Accountabilities for Data Quality Management	ACIS 2007 Proceedings	Literature Review	CDQM
2009	Batini, Carlo; Capiello, Cinzia; Ciara Francalanci, Ciara and Maurino, Andrea	Methodologies for Data Quality Assessment and Improvement	ACM Computing Surveys	Literature review	Data Quality
2010	Lucas, Ana	Towards Corporate Data Quality Management	Portuguese Journal of Management Studies	Empirical study: case study	Corporate Data Quality Management
2012	Otto, Boris; Huner. Kai M.; Österle, Hubert	Toward a functional reference model for master data quality management	Inf. Syst E-bus Manage	Empirical study: case study	Data Quality Management, Masterdata Management

Table 15. Articles used in subsection 4.4.1

Lucas (2010) defines *corporate data philosophy* according to corporate business philosophy and how it should promote the idea that the data is an asset that creates value for the business. *Corporate data policy* aligns with corporate data philosophy and defines guidelines for governing data in terms of sharing and using the data and maximizing the value of the data. It should also cover the following interrelated categories: data quality dimensions for every data set, data assets catalog, data sharing, availability and accessibility, data architecture, data security and appropriate use and data planning. *Corporate data strategy* "is a long term plan of action designed to achieve the directions prescribed by CDP in line with Corporate Business Strategy" (Lucas, 2010, p. 177).

Data governance can be considered as the practice of practical management over the data assets. Lucas (2010) defines four activities for data governance, first one being management and leadership over the data assets. The second activity is the planning and carrying out the data quality strategy in line with the corporate data philosophy. The third goal is to provide the needed resources and structures that are needed to carry out the planned strategy. The fourth activity is getting people involved with the corporate data strategy and its goals.

An organization also has to define *data quality management* for itself and its employees. Lucas (2010, p. 178) defines it "as a set of coordinated activities to direct and control data quality activities in an organization," whereas Otto et al. (2007, p. 918) define it "as quality-oriented data management, i.e., data management focussing on collection, organisation, storage, processing, and presentation of high-quality data."

Data owner is a person or a unit who has the responsibility over the data set. (Lucas, 2010) *Data quality methodology* is a set of phases, activities, techniques and tools to achieve the goals set in data quality strategy (Lucas, 2010). This was discussed in section 4.1.3 and 4.1.4 which presented the models of Batini et al. (2009) and Wang (1998). *Data quality techniques* are either data driven or process driven. The data driven techniques concentrate on certain problems in the data quality and the process driven techniques concentrate on improving the processes that create the data (Lucas, 2010). *Data quality tools* are specific software products that carry out certain data quality techniques to achieve the desired data

quality level (Lucas, 2010).

Master data informs about the most important business entities of a corporation, for example operations, suppliers, customers, products, employees, and assets (Otto, Hüner, and Österle, 2012). Due to the large amount of undefined data, various academics have chosen to concentrate on the management of master data. Due to the continuous interchangeable use of various terms, this thesis uses models and concepts introduced for *master data quality management* for overall corporate data quality management model. Master data quality management is concentrated on managing essential organizational data (Lucas, 2010).

According to Lucas (2010), *data stewards* are persons responsible of certain data sets. Wende (2007) promotes the model of dual management of data, one steward for business and one for technology. She points out five important roles in data quality management and sets them in a RACI table. The five roles are: executive sponsor, data quality board, chief steward, business steward, and technical steward.

Defining data quality philosophy, policy and strategy are especially important in creating a more coherent and value-creating corporate data quality management model. Involving both technical and business employees in planning the model helps integrating both technical and business goals to the corporate data quality management. Creating a philosophy and strategy is also the first step to a more strategic approach craved by Otto et al. (2007), which in turn is a step towards more preventive data quality management system. This ties the theory presented to the section 4.3 that explores the current problems in data quality management.

4.4.2 Data Quality Management Model

This subsection introduces the functional reference model for master data quality management by Otto, Hüner and Österle (2012). The entire section is based on this research. Search terms used in this subsection are *master data* and *master data management.*

Articles used in this subsection are listed in Table 16.

Year	Authors	Title	Journal	Approach	Focus Area
2012	Otto, Boris; Huner. Kai M.; Österle, Hubert	Toward a functional reference model for master data quality management	Inf. Syst E-bus Manage	Empirical study: case study	Data Quality Management, Masterdata Management

Table 16. Articles used in subsection 4.4.2

Most studies on data quality management discovered in the systematic literature review, concentrate on only one aspect or function of it. Lucas (2010) emphasizes the need for a clear model to help in the review. This model provides a structure that is to be used when reviewing the literature on corporate data management. Further theory is added to the listed function groups. The model used offers one coherent organizational model for managing data quality and how to link separate organizational sections to support it.

The model divides corporate data quality management into six function groups that are listed in the far left column in the gray-coloured boxes. These function groups include concrete functions in the corporate data quality management process. The functions are listed in the white boxes right from the gray function group column. The function groups are briefly described in this section in Figure 7. One subsection for each function group follows.

Master Data Lifecycle Management	Data Creation	Data Maintenance	Data Deactivation	Data Archiving
Metadata Management and Master Data Modeling	Data Modeling	Model Analysis	Metadata Management	
Data Quality Assurance	Data Analysis	Data Enrichment	Data Cleansing	
Master Data Integration	Data Import	Data Transformation	Data Export	
Cross Functions	Automation	Reports	Search	Workflow Management
Administration	Data History Management	User Management		

Figure 7. Master data quality management model developed by Otto et al (2012, p. 405).

The first function group *Master Data Lifecycle Management* includes all activities data users or data managers conduct with master data during its lifecycle (Otto et al., 2012). In the second function group *Metadata Management and Master Data Modeling,* metadata defines the properties of data and the meaning of data. This is continued with "metadata comprises all the information necessary for efficient management and effective usage of master data. According to the definition above, *master data modeling* means creation of technical metadata (data types, relationship multiplicities, for example)" (Otto et al., 2012, p. 405).

The third function group, *Data Quality Assurance* "comprises functions for preventive (i.e. to prevent potential, future data defects) and reactive (i.e. to repair data defects that have occurred) maintenance and improvement of the quality of master data. The three functions

[*Data Analysis, Data Enrichment* and *Data Cleansing*] comprise sub-functions for identification of data defects and measuring data quality (*Data Analysis*), for improving data quality by comparison with and integration of external reference data (*Data Enrichment*) and for repair of data defects identified (*Data Cleansing*)" (Otto et al., 2012, p. 405).

The fourth function group *Master Data Integration* "comprises functions supporting transfer (import and export) and structural transformation (e.g. consolidation of fields or tables) of master data" (Otto et al., 2012, p. 405). The fifth function group *Cross Functions* are "functions that cannot be assigned to one of the other groups. Sub-functions under the function *Automation* do not provide additional functionality but offer support for being able to efficiently use other functions by making them machine processable" (Otto et al., 2012, p. 405). The sixth function group *Administration* function "comprises functions for user administration and the tracing of changes and modifications made" (Otto et al., 2012, p. 405).

This research explores the data quality management through this model. These presented six function groups form the core model to which further theory is added to answer the defined research question: How should a corporation manage its data quality?

4.4.3 Data Lifecycle Management

This section assesses the further literature on the first function group in the model is master data lifecycle management which can be dubbed as data lifecycle management. Otto et al. (2012) define data lifecycle management as a process that begins from the creation of the data and ends at the deletion and archiving of the data. The four functions in this group are *Data Creation, Data Maintenance, Data Deactivation* and *Data Archiving.* Search terms used in this subsection are *data quality, corporate data quality, corporate data quality management, master data* and *master data management.*

Articles used in this subsection are listed in Table 17.

Year	Authors	Title	Journal	Approach	Focus Area
1998	Wang, Richard Y	A Product Perspective on Total Data Quality Management	Communications of the ACM	Empirical study: survey research	Data Quality

2009	Hüner, Kai M.; Martin Ofner, Martin and Otto, Boris	Towards a Maturity Model for Corporate Data Quality Management	SAC 09 Proceedings of the 2009 ACM Symposium on Applied Computing (2009)	Empirical study: case study	Corporate Data Quality Management
2012	Ofner, Martin H; Otto, Boris; Österle, Hubert	Integrating a data quality perspective into business process management	Business Process Management Journal	Empirical study: case study	Data Quality, Business Process Management
2012	Otto, Boris; Huner. Kai M.; Österle, Hubert	Toward a functional reference model for master data quality management	Inf. Syst E-bus Manage	Empirical study: case study	Data Quality Management, Masterdata Management
2013	Elbireer, Ali; Le Chasseur, Julie; Jackson, Brooks	Improving laboratory data entry quality using Six Sigma	International Journal of Health Care Quality Assurance	Empirical study: case study	Data Quality
2013	Ofner, Martin H; Straub, Kevin; Otto, Boris and Österle, Hubert	Management of the masterdata lifecycle: a framework for analysis	Journal of Enterprise Information Management	Empirical study: case study	Masterdata management
2015	Huang, Hong	Domain knowledge and data quality perceptions in genome curation work	Journal of Documentation	Empirical study: survey research	Data Quality

Table 17. Articles used in subsection 4.4.3

The data lifecycle consists of the four functions: *Data Creation, Data Maintenance, Data Deactivation* and *Data Archiving. Data Creation* is the function in which the data is created. The function includes a data entry process in which the data is inserted into the system. *Data Maintenance* is the function that may include possible plausibility checks in order to validate the data. *Data Deactivation* is the function in which the data is deactivated in the system and *Data Archiving* is the function in which the data is archived so that in can be reviewed in the future (Otto et al., 2012).

The process of creation, maintenance, deactivation and archiving of the data affects greatly the quality of the data, so planning and implementing correct set of processes for data lifecycle management is crucial for a successful data quality model. Ofner, Straub, Otto and Oesterle (2013) note that the area of data lifecycle management suffers from two major problems: the lack of non-preventive coordination of the data lifecycle and the lack of business involvement.

Ofner et al. (2013) develop the data lifecycle management model further. They consider that one of the major problems in the current data lifecycle management is resulted from the current system of data lifecycle management in which the data goes through its lifecycle of *Data Creation, Data Maintenance, Data Deactivation* and *Data Archiving* without coordination. The data is more likely to be used by separate business domains that have their own, very specific information needs which often results in uncoordinated processing of the data resources. The coordinated processing of the data resources would be more preventive approach for data lifecycle management. They continue their study with the notion that despite the importance of a functioning data lifecycle model, the existing models are very generic. (Ofner et al., 2013)

Ofner et al. (2013) also emphasize the theory of Wang (1998) stating that companies concentrate too much on the IT systems which leads to an organization that considers data as a by-product of the system. This is further supported by Hüner et al. (2009). The problem with this attitude is that it gives the data the position of a technical liability and not the business asset that creates value. This links this section to section 4.3 that identified the lack of business aspect as a crucial issue in data quality management.

Ofner et al. (2013) and Elbireer, Le Chasseur and Jackson (2013) emphasize that the data should be treated as a product that is used for the business purposes. This idea is continued by the notion that the data is always a product that comes from a specific process. This links this section to section 4.2 that emphasized that data is always a product of a process. Employees responsible for data creation must understand the business needs of the data consumers and build a process that supports the creation of this data.

Ofner et al. (2013) have developed a master data quality lifecycle map which divides the function of master data lifecycle management into four task groups: data portfolio, data and system design, data supply chain, and data support. These tasks are linked to business components that have been divided into three levels: direct, control and execute. The direction is taken from the master data strategy. Master data quality lifecycle map brings clarity to the complex problem of managing data lifecycle in a multi-functional organization. It brings the business aspect to the data lifecycle management. The map can be seen as a checklist for

organizational jobs that have to be performed in order to manage the data lifecycle. The map is illustrated in Figure 8.

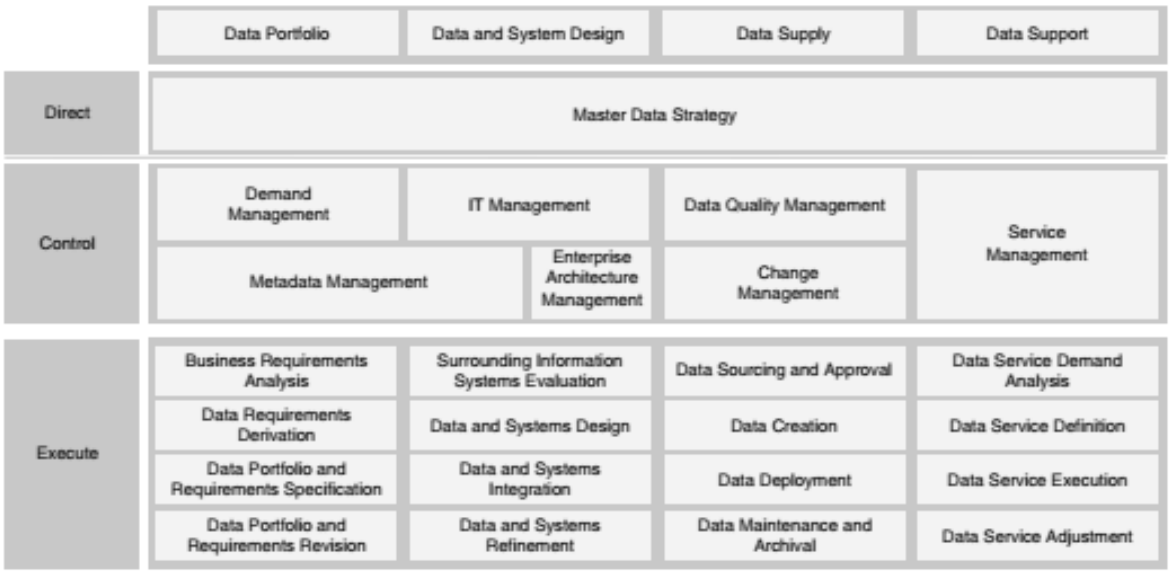

Figure 8. Master Data Quality Lifecycle map (Ofner et al., 2013, p. 480).

Task group data portfolio links the business requirements to the master data lifecycle management by specifying the business requirements for the data and building the data portfolio on the basis of the requirements. Two important tasks of this task group are to define the scope of the master data quality management and "and develop guidelines on how the master data portfolio and data requirements need to be specified and documented" (Ofner et al., 2013, p. 479). The third task is to supervise the specifications and maintenance of the portfolio requirements (Ofner et al., 2013).

Task group data systems and design links the business requirements to the technical master data architecture. This is done by developing guidelines for the master data architecture, monitoring "implementation or enhancement of required applications and databases" (Ofner et al., 2013, p. 479-480), and maintaining the technical metadata. This group also includes the design work of master data models, processes, architectures, and IT systems that guarantee the readiness of the IT systems (Ofner et al., 2013).

Task group data supply chain links the business requirements to the production and maintenance processes of the data assets. This is done by developing instructions for the creation and maintenance of master data and monitor its quality. The group also includes managing requests, collecting data components, creating and deploying master data (Ofner et al., 2013). This links the model to section 4.2 and the data production processes.

Task group data support links the business requirements to the data support services. This is done by specifying the required data support services, monitor the incident requests and problem resolution times, and "offer master data services and post-consultancy support required by data users" (Ofner et al., 2013, p. 480). The offered master data services may include project support functions and mass data changes (Ofner et al., 2013).

Ofner et al. (2013) and Otto et al. (2012) emphasize the management of the master data lifecycle for which the master data lifecycle map brings clarity and preventive strategy which is also one of the identified problems in data quality management. Another point they explore is the data creation process itself. Ofner et al. (2013) emphasize that different business functions follow their own needs in creating, using and deleting the data which leads to a very uncoordinated and expensive situation in which the full business potential of the data cannot be used.

The team of Ofner, Otto and Österle (2012) continue developing the data creation process by exploring the preventive possibility of integrating the data quality perspective into business processes by planning data producing and using processes so that they produce quality data for the business purposes that can be used to create value in the business. Ofner et al. (2012) describe a set of six requirements for the process of integrating data quality perspective to the business process management. The requirements can be seen as guidelines for the integration. The requirements developed by Ofner et al. (2012) are listed below.

1. The organization should provide definitions and guidelines about how to measure the data quality in designing the business process. This would create a new success factor in the design process.
2. The modelling language of the integrating method should reuse and extend the already established concepts of business process management as extensively as possible. This

relates the new concepts to the old, established ones.

3. The data quality evaluation must be provided on various levels of abstraction because there are various data consumers who have different information needs.
4. The organization needs to provide concepts on how to emphasize the business impact and value of the data quality project.
5. The organization must confirm that the business process models are able to support the IT implementation.
6. The organization must provide relevant data context and enable different representation that is associated to the surrounding action system.

In managing the data lifecycle and in bringing the data quality aspect to business processes, the aspect of employee training needs to be explored. The importance of training the employees must be emphasized to have an appropriate portfolio of data quality skills that consists of adaptive skills, interpretative skills, data quality literacy skills and technical skills. Adaptive skills include skills to turn subjective data quality user requirements into objective technical specification. Other skill for the responsible employee is the ability to analyze and develop the process to achieve a desired level of data quality. This should also include a skill to establish and maintain organizational policies that would ensure a certain level of data quality. There should also be skills for managing information overload due to the large volume of data. All this would also require change management skills and cost/benefit calculation skills to find out the benefits of the initative. (Huang, 2015)

Interpretative skills include skills that support the reactive aspect of data quality management. It includes skills to detect and correct errors in databases. There is also a need to use relevant software to carry out this function. (Huang, 2015)

Data quality literacy skills include skills that enable the relevant employees to assess and measure data quality. The skill set includes an understanding of the relevant data quality dimensions and measurements and the skill to use them in order to assess the data. This skill set also includes a thorough understanding of the modern data quality issues and their possible effects to the organization. (Huang, 2015)

The technical skills include skills for formal audit of the data quality. It also includes a set of skills to assess and use the data including statistical techniques to make sense out of the data, analytical models, data mining abilities to analyze data in a data warehouse, and skills to integrate databases into a data warehouse. The final skill is SQL skill to work on the data. (Huang, 2015)

This subsection explores the studies of data lifecycle management of a corporation. The function group was explored through four functions *Data Creation, Data Maintenance, Data Deactivation* and *Data Archiving.* The theory of data lifecycle management was further developed by Ofner et al. (2013) who developed a data quality lifecycle map that is a tool of data lifecycle management that clarifies the data creation process view and adds business perspective to data lifecycle management. After this, they continue developing idea of bringing data quality perspective to business processes through six requirements that can be seen as a preventive measure in data quality management. Huang (2015) continues the data lifecycle studies by exploring the training needs of the employees.

Combined these theories form a set of functions to be conducted in order to improve the data lifecycle management. The also tackle the problems and issues explored in the previous sections. The models of data creation process improvement involve the process aspect more to the data quality management which ties the theory to section 4.2. The models also prevent data quality problems and add business aspect to the data quality management.

4.4.4 Metadata Management and Master Data Modeling

This subsection concentrates on the second function group, managing and modeling the metadata of the data. According to Otto et al. (2012) the "metadata comprises all the information necessary for efficient management and effective usage of master data" (Otto et al. 2012, p. 405). This function group consists of *Data Modeling, Model Analysis* and *Metadata Management*. Search terms used in this subsection are *corporate data quality, corporate data quality management, master data* and *master data management.*

Articles used in this subsection are listed in Table 18.

Year	Authors	Title	Journal	Approach	Focus Area
2011	Hüner, Kai M.; Otto, Boris and Österle, Hubert	Collaborative Management of Business Metadata	International Journal of Information Management	Empirical study: case study	Data Quality Management, Metadata Management
2012	Otto, Boris; Huner. Kai M.; Österle, Hubert	Toward a functional reference model for master data quality management	Inf. Syst E-bus Manage	Empirical study: case study	Data Quality Management, Masterdata Management
2016	Aljumaili, Mustafa; Karim, Ramin and Tretten, Phillip	Metadata-based data quality assessment	VINE Journal of Information and Knowledge Systems	Narrative Literature review	Metadata, data quality

Table 18. Articles used in subsection 4.4.4

Aljumaili, Karim, and Tretten (2016) add the definition of metadata. They define three main types of metadata: descriptive metadata (title, abstract, author, and keywords) for discovery and identification, structural metadata for indicating how many objects are combined, and administrative metadata for managing the resource. They note that the quality of metadata is an essential element in the data because it is the basis of database management. They also note that it is one of the earliest data quality research points.

Otto et al. (2012) include three functions to the function group: *Data Modeling, Model Analysis* and *Metadata Management*. Data Modelling includes data model editing and classification in which the form of data is modelled to comply with the business needs of the company. In the Model Analysis the data is further analysed. Metadata Management consists of functions that are meant to transform the metadata into a manageable form.

Aljumaili et al. (2016) develop further *Data Modeling* by noting that metadata records the basic data about the data (for example users, schemas, tables, columns and indexes) which help the users locate and withdraw data. Because of this metadata should have one modelled format inside an organization in order to achieve clarity. The metadata can be controlled through specific constraints that are certain rules for the metadata developed in *Model Analysis*. They list not null, unique, primary key, foreign key and check to be the most common constraints. They suggest four *Model Analysis* activities in order to analyse the quality of metadata: primary key check, foreign key check, check constraint and nullability check.

Another *Data Modeling* function recommendation is to check primary keys for

containing unique values and not containing null values. There should also be a special emphasis on checking whether there is a primary key, since not having one makes database management difficult. They separate foreign key constraint from primary key constraint with the notion that the constraints of primary key do not apply to the constraints of foreign key. (Aljumaili et al., 2016)

Check constraints strengthen domain integrity by defining the range of admissible values. They are important in defining and specifying the metadata. Null value is a missing value that is known to exist in the real world but is unknown. It is crucial to check this value to understand the missing parts of metadata. (Aljumaili et al., 2016)

Hüner, Otto, and Österle (2011) bring the organizational aspect by concentrating on *Metadata Management* on the business objects, business metadata. They recognize a problem that the traditional, centrally organized metadata managers are not the most optimal way of organizing the metadata management when "the knowledge needed to create consistent business metadata is spread across various departments, divisions or lines of business" (Hüner et al., 2011, p. 367). They call for collaborative management for the business metadata in which several business and technical experts would provide their expertise. This could be done by constructing a business metadata repository that could be wiki-based. The group has developed a set of requirements for the business metadata repository in Table 19 (Hüner et al., 2011).

Requirement	Description
R01. Cross-links between metadata	In dictionaries, terms are usually defined with the help of cross-links to other terms, glossaries, or pictures. Likewise, the business metadata repository should contain such cross-links in order to avoid metadata redundancy and allow for efficient navigation.
R02. Standardized metadata format	The metadata kept in the business metadata repository should be highly visible, i.e. it should be available and used by as many enterprise systems as possible, thereby fostering the willingness to take part in the collaboration efforts among as many users as possible. A standardized metadata format would facilitate easy metadata import and export.
R03. Flexible user interface	The business metadata repository's user interface should offer options for flexible configuration (i.e. different roles for different jobs) in order to accomplish effective collaboration, including as many users, departments and divisions as possible.
R04. Integration with portals and intranets	A frequent constraint stipulated by a company's IT strategy is to reduce the amount of systems visible to users and to provide a single front end (e.g. a portal or intranet). For that reason it should be possible to integrate the business metadata repository's front end with existing information systems.
R05. Multilanguage capability of software and metadata	In order to accomplish effective collaboration (cf. R03), metadata must be available in different languages. Also, the business metadata repository's user interface should be multilanguage capable.
R06. Support of usage and maintenance processes	Usage and maintenance processes should be clearly specified and documented (e.g. searching metadata within the business metadata repository, workflows for updating metadata, or reporting defects) in order to make sure the business metadata repository is efficiently used.
R07. Change history	A metadata change history should be available at any time in order to be able to quickly identify modifications made, correct errors, and recover previous states.
R08. Flexible metadata structure	Over time, metadata is likely to require modification (regarding both meaning and structure), caused by e.g. the need to describe new data attributes or to implement additional enterprise systems that use the data. In order to ensure high metadata quality (i.e. up to date and accurate), it should be easy to make modifications.
R09. Single-sign-on support	The business metadata repository should be capable of integrating existing modes for authentication (i.e. no separate authentication should be required).

Table 19. "Requirements to be met by business metadata repository" (Hüner et al., 2011, p.

369).

This subsection explored the metadata management and masterdata modelling through three functions *Data Modeling, Model Analysis* and *Metadata Management*. These concepts are further developed by the theories of Aljumaili et al (2016) who studied the formatting of metadata. When these studies are combined, they form a set of activities and best practices to manage the metadata.

4.4.5 Data Quality Assurance

This subsection concentrates on the third function group in the reference model is data quality assurance which "comprises functions for preventive (i.e. to prevent potential, future data defects) and reactive (i.e. to repair data defects that have occurred) maintenance and improvement of the quality of master data. The three functions comprise sub-functions for identification of data defects and measuring data quality (*Data Analysis*), for improving data quality by comparison with and integration of external reference data (*Data Enrichment*) and for repair of data defects identified (*Data Cleansing*)" (Otto et al., 2012, p. 405). Search terms used in this subsection are *corporate data quality, corporate data quality management, master data* and *master data management.*

Articles used in this subsection are listed in Table 20.

Year	Authors	Title	Journal	Approach	Focus Area
2011	Derby, Dustin C.; Haan, Andrea and Wood, Kurt	Data quality assurance: an analysis of patient non-response	International Journal of Health Care Quality Assurance	Empirical study: survey research	Data Quality Assurance
2012	Elgammal, Amal, Oktay Turetken, and Van Den Heuvel, Willem-Jan	Using Patterns for the Analysis and Resolution of Compliance Violations	International Journal of Cooperative Information Systems	Narrative Literature Review	Regulatorly Compliance
2012	Otto, Boris; Huner. Kai M.; Österle, Hubert	Toward a functional reference model for master data quality management	Inf. Syst E-bus Manage	Empirical study: case study	Data Quality Management, Masterdata Management

Table 20. Articles used in subsection 4.4.5

Derby, Haan, and Wood (2011) develop data quality assurance further by defining it as "activities of profiling and analyzing data for the purposes of understanding the potential causes of data artifacts (response rates, reliability, validity), and anomalies" (Derby, Haan, and Wood, 2011, p. 199). They also note that the objective of the data quality assurance and the three subfunctions of it should be determining the current state of the data quality through various different evaluation techniques. The theory for defining a desired level of data quality can be found in section 4.2.3 which explores the theoretical standards of data quality. Different departments have to define their standards for the data so that it is fit to be used (Otto, Hüner, and Österle, 2012).

An important tool in assuring data quality is to develop a functioning set of compliance controls around the data management practices. According to Elgammal, Turetken and Van Den Heuvel (2012) describe compliance as a process that aligns business and IT processes to comply with relevant regulation and best practices. Controls can also develop the data quality management into more preventive direction. The control framework can be found in widely-used control frameworks, such as ISO 9001 (Elgammal et al., 2012).

This subsection explored the data quality assurance through functions *Data Analysis, Data Enrichment* and *Data Cleansing.* These functions were further developed by Derby et al. (2011) who also link the topic to data quality assessment in section 4.2. Elgammal et al. (2012) continue developing a theory on controls that form effective means to prevent data quality problems.

4.4.6 Data Integration

This subsection concentrates on the fourth function group listed by Otto et al (2012), master data integration consisting of *Data Import, Data Transformation* and *Data Export*. It "comprises functions supporting transfer (import and export) and structural transformation (e.g. consolidation of fields or tables) of master data" (Otto et al., 2012, p. 405). Data import and export formats are emphasized in the import and export functions and the data type conversion is emphasized in the data transformation (Otto et al., 2012). This subsection explores the Data Integration function by exploring the practical application of data integration, data warehousing. After this the data quality of ERP is studied. Search terms used in this

subsection are *data quality, corporate data quality, corporate data quality management, master data* and *master data management.*

Articles used in this subsection are listed in Table 21.

Year	Authors	Title	Journal	Approach	Focus Area
1993	John, S.A.	Data integration in GIS - the question of data quality	Aslib Proceedings	Empirical study: case study	Data Integration
1996	Wand, Yair and Wang, Richard Y.	Anchoring Data Quality Dimensions in Ontological Foundations Foundations	Communications of the ACM	Narrative Literature review	Data Quality
1997	Hurley, Margaret A. and Harris, Rod	Facilitating corporate knowledge: building the data warehouse	Information Management & Computer Security	Narrative Literature review	Data Warehouse
1998	Cheng, Paul S. and Chang, Pintsang	Transforming corporate information into value through data warehousing and data mining	Aslib Proceedings	Empirical study: case study	Data Warehouse
2000	Ma, Catherine; Chou, David C. ; Yen, David C.	Data warehousing, technology and assessment	Industrial Management & Data Systems	Narrative Literature review	Data Warehouse
2002	Xu, Hongjiang; Nord, Jeretta Horn; Brown, Noel; Nord, G. Daryl	Data quality issues in implementing ERP	Industrial Management & Data Systems	Empirical study: case study	Information Management, Database Management, Quality
2007	Scarisbrick-Hauser, AnneMarie, and Rouse, Christina	The Whole Truth and Nothing but the Truth? The Role of Data Quality Today	Direct Marketing: An International Journal	Narrative Literature review	Data Quality
2009	Haug, Anders; Stentoft, Jan Arlbjorn; Pedersen, Anne	A Classification Model of ERP System Data Quality	Industrial Management & Data Systems	Empirical study: case study	ERP Data Quality
2012	Otto, Boris; Huner. Kai M.; Österle, Hubert	Toward a functional reference model for master data quality management	Inf. Syst E-bus Manage	Empirical study: case study	Data Quality Management, Masterdata Management
2014	Glowalla, Paul; Sunyaev, Paul	ERP System Fit - an Explorative Task and Data Quality Perspective	Journal of Enterprise Information Management	Empirical study: in-depth interviews	ERP Data Quality
2014	Kwon, Ohbyung; Namyeon Lee,	Data Quality Management, Data Usage	International Journal of	Empirical study: survey research	Data Quality Management,

	Namyeon and Shin, Bongsik	Experience and Acquisition Intention of Big Data Analytics	Information Management		Masterdata Management

Table 21. Articles used in subsection 4.4.6

John (1993) discusses that the traditional problem with integrated databases is that different users create datasets that differ from each other by accuracy, standards and manner they are created. “Analysis carried out on an integrated database can ultimately be as accurate as the quality of the ‘worst’ data set included” (John, 1993, p. 109). This subsection explores the positive aspects of data warehousing and ERP systems to data quality.

Both Kwon, Lee and Shin (2014) and Scarisbrick-Hauser and Rouse (2007) note that the data is created in different departments of different organizations that causes the data quality to vary. Cheng and Chang (1998) note that organizations struggle with the variety of data imported from different departments. They also call for a change from considering data as a technical by-product, and giving it the status of a strategic asset which links this to the identified problems in data quality management. They call for data warehouse to integrate the data. "A data warehouse is a subject oriented, integrated, non-volatile, time-variant collection of data organized to support management needs" (Cheng and Chang, 1998, p. 109). The purpose of the warehouse is to integrate data from various sources and make it available for the end users (Ma, Chou, and Yen, 2000). Hurley and Harris emphasize that most data warehousing projects do not fail because of the technological issues, but because of the organizational issues (Hurley and Harris, 1997).

The data warehouse has to support the business decisions and note that the problem has been the lack of business involvement in the process. They recognize two types of managers needing information: operational managers and business analysts/executives. The operational managers are responsible for the daily routines and management of the company, whereas the business analysts are in charge of the more long-term strategy of the company. (Cheng and Chang, 1998)

Haug, Arlbjorn, and Pedersen (2009) continue the idea of data integration with their studies on ERP systems. According to them, the current problem with enterprise data is its variety of sources (departments, geographical locations) which leads into the formation of data

silos. When the data is available throughout the organization, the result is often inefficiencies in the operations. They also note the link between data quality and ERP system by noting that the success of the ERP project is often determined by the quality of data (Haug et al., 2009).

They continue that the lack of data quality dimensions for ERP project and develop a data quality classification model which allows employees responsible to evaluate the data quality in the ERP system. They have used the earlier works of Wang, Strong and Shank and identified three data quality categories to evaluate the data quality: intrinsic data quality dimensions, data accessibility dimensions and data usefulness dimension. (Haug et al., 2009)

The intrinsic data quality dimension (adopted from Wand and Wang, 1996) measures the completeness, unambiguousness, meaningfulness, and correctness of the data. The data accessibility dimensions measure the access rights to the system. According to Haug et al. (2009) these two categories are the easiest to measure due to their measurable nature compared to the very relevant usefulness dimension. They develop a two-dimensioned matrix to evaluate the ERP data which evaluates the data on the basis of intrinsic and accessibility dimensions and rates the data from 1 to 4.

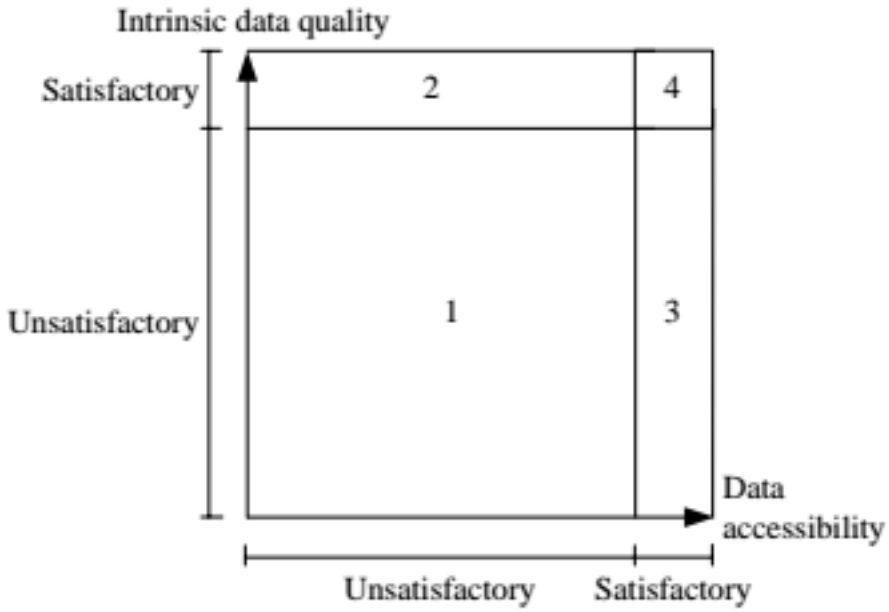

Figure 9. ERP data quality evaluation model (Haug et al., 2009, p. 1059).

Area 1 in the model is unsatisfactory at both dimensions. The second area has a level of intrinsic quality, but it cannot be accessed by every related employee in the organization, whereas the third area of the model has a poor level in the intrinsic value, but can be accessed by every related employee. The fourth area is the ultimate goal of the ERP data quality project. (Haug et al., 2009)

The intrinsic aspect of the data quality is often emphasized and that the modern ERP data quality management is lacking in contextual data quality. This means that the current systems provides data that has good quality for one task, or context, but does not work well in other contexts. They note that the modern ERP data has to have a contextual aspect so that the employees have to be able to use the data in various different contexts. (Glowalla and Sunyaev, 2014)

Xu, Nord, Brown and Nord (2002) note that the issues of data quality are apparent already in the system implementation phase and that the system is going to phase difficulties if the data in it is faulted and has poor quality. They promote employee training, the support of the top management, communication, change management, employee relations and data quality controls in managing the implementation of ERP system.

This subsection explored a practical data quality issue of data being formed in silos which causes the data to be created without coordination, common standards and business aspect which links this to the data quality problems listed in section 4.3. An important practical tool in preventing these issues is to create integrated data warehouses and ERP solutions with which the data can be used better. This subsection explored the tools that can be used to improve the quality of the data in the systems.

4.4.7 Cross Functions

This subsection concentrates on the fourth function group, cross functions. The function group consists of "functions that cannot be assigned to one of the other groups" (Otto et al., 2012, p. 405). This function group can be seen as a group of support functions for different purposes. Otto et al. (2012) describe four subgroups in this function group: *Automation, Reports, Search*

and Workflow Management. Search terms used: *data quality, corporate data quality, corporate data quality management, master data* and *master data management.*

Articles used in this subsection are listed in Table 22.

Year	Authors	Title	Journal	Approach	Focus Area
1998	Cheng, Paul S. and Chang, Pintsang	Transforming corporate information into value through data warehousing and data mining	Aslib Proceedings	Empirical study: case study	Data Warehouse
2009	Haug, Anders; Stentoft, Jan Arlbjorn; Pedersen, Anne	A Classification Model of ERP System Data Quality	Industrial Management & Data Systems	Empirical study: case study	ERP Data Quality
2012	Elgammal, Amal, Oktay Turetken, and Van Den Heuvel, Willem-Jan	Using Patterns for the Analysis and Resolution of Compliance Violations	International Journal of Cooperative Information Systems	Narrative Literature review	Regulatorly Compliance
2012	Ofner, Martin H; Otto, Boris; Österle, Hubert	Integrating a data quality perspective into business process management	Business Process Management Journal	Empirical study: case study	Data Quality, Business Process Management
2012	Otto, Boris; Huner. Kai M.; Österle, Hubert	Toward a functional reference model for master data quality management	Inf. Syst E-bus Manage	Empirical study: case study	Data Quality Management, Masterdata Management

Table 22. Articles used in subsection 4.4.7

According to Otto, et al. (2012), automation merely eases the process of other functions and makes them machine processable. The subfunction consists of automated enrichment, automated export, automated import, cross-function automation and push and pull mechanisms. Automated enrichment enables automatic comparing of data with external reference data.

Automated export and import functions enable the automated exchange of data between a test system and transaction systems (Otto et al., 2012). “Cross-function automation allows automated execution of various, linked functions in a certain sequence” (Otto et al., 2012, p. 421). Push and pull mechanisms enable to use push- and pull mechanisms for automated data import and export. This is an important practical add to the Data Integration discussed in

subsection 4.3.6. Automation is an integral supporting function in data warehouse management explored by Cheng and Chang (1998) and ERP systems explored by Haug et al. (2009) because both functions create value partly through automation.

Reporting function produces reports for various stakeholders involved with the project. The reports include data quality reports, usage statistics, job monitoring and audit support. The search function allows users to carry out various searches which include dynamic value search, free search and fuzzy search. Both functions clarify and ease the management and monitoring of the system (Otto et al., 2012). Comprehensive reporting is also important for effective compliance discussed by Elgammal et al. (2012) in subsection 4.3.5. Effective reporting can be considered to be a basic requirement for data quality compliance because the compliance procedures are based on the reporting material.

Workflow management allows bundling of activities, graphical workflow modelling and the creation and maintenance of workflows. The first subgroup enables the bundling of several activities within a single MDM workflow. Graphical workflow modeling enables the modeling of the workflows with different graphical symbols. The creation and maintenance of workflows allows the management activity sequences of different processes and departments (Otto et al. 2012). Bundling of activities, graphical workflow modelling and the creation and maintenance of workflows are important practical activities in integrating data quality into business processes that has been explored Ofner et al. (2012) in subsection 4.4.3.

Cross Functions can be seen as tasks that cannot be classified to other function groups but are still necessary for the data quality management. The tasks vary from administrative tasks to automated tasks. They are both reactive and preventive.

4.4.8 Administration

This subsection concentrates on the sixth function group administration. It "comprises functions for user administration and the tracing of changes and modifications made" (Otto et al., 2012, p. 405). The function group comprises functions of *Data History Management* and *User Management* (Otto et al., 2012). Various researchers have been studying who should do what in the process of data quality management. The question is considered as one of the most

important questions in involving the business needs to the data quality management (Wende, 2007). Search terms used in this subsection are *corporate data quality* and *corporate data quality management.*

Articles used in this subsection are listed in Table 23.

Year	Authors	Title	Journal	Approach	Focus Area
2007	Wende, Kristin	A Model for Data Governance - Organising Accountabilities for Data Quality Management	ACIS 2007 Proceedings	Narrative Literature review	CDQM
2012	Otto, Boris; Huner. Kai M.; Österle, Hubert	Toward a functional reference model for master data quality management	Inf. Syst E-bus Manage	Empirical study: case study	Data Quality Management, Masterdata Management

Table 23. Articles used in subsection 4.4.8

Wende (2007) develops the administration theory with the model of dual management of data in which the responsibility over data has been divided to specific data stewards, one steward for business and one for technology. The aim of the model is to involve business aspect to the data quality management and make data quality management more preventive. She points out five important roles in data quality management and sets them in a RACI table in Table 24. The five roles are: executive sponsor, data quality board, chief steward, business steward, and technical steward.

"The executive sponsor is a member of the top management, such as the CEO, CFO or CIO. Besides supporting DQ initiatives and data governance, he or she provides sponsorship, strategic direction, funding, advocacy and oversight for data quality management" (Wende, 2007, p. 421). She emphasizes that the support from the top management level is essential considering the cross-organizational nature of the project.

The task of the data quality board is to plan and define the data governance model for the entire organization. The board makes strategic goals and align them with the business mission and objectives of the company. "More particularly, it develops and directs corporate-wide standards, rules, policies, processes, and guidelines to ensure the ongoing improvement of DQ" (Wende, 2007, p. 421). The board then "provides mechanisms for coordination,

communication, information sharing, prioritisation, and conflict resolution. The DQ board is usually chaired by the chief steward" (Wende, 2007, p.421).

"The main task of the chief steward is to put the board's decisions into practice. He or she enforces the adoption of standards, helps establish DQ metrics and targets, and ensures that regulatory, privacy and information sharing policies are followed. In addition, the chief steward staffs and supervises all data stewards, but also helps them to enforce their mandates" (Wende, 2007, p. 421).

Business data stewards operate with the business representatives. Their task is to prepare the business requirements documentation and studies the possible effects of new business requirements on data quality and vice versa (Wende, 2007). "Usually, one business data steward is assigned either per business unit, per main business process or per main data type. For their area of responsibility, the business data stewards detail the corporate-wide DQ standards and policies brought up by the board. This may involve creating business rules for data, developing data vocabularies, and maintaining and publishing DQ metrics" (Wende, 2007, p. 421). Their task is to communicate their knowledge to the data quality board and make recommendations considering standards and policies derived from the business requirements (Wende, 2007).

The task of the technical data stewards is to represent the the data in IT systems. The technical data steward can be appointed either per business unit or department or per IT system (Wende, 2007). Their task is to "provide standardised data element definitions and formats for their area of responsibility and focus on technical metadata. In addition, technical stewards profile source system details and data flows between systems. They communicate IT-related requirements to the DQ board" (Wende, 2007, p. 421).

Roles / Decision Areas	Executive Sponsor	Data Governance Council	Chief Steward	Business Data Steward	Technical Data Steward	...
Plan data quality initiatives	A	R	C	I	I	
Establish a data quality review process	I	A	R	C	C	
Define data producing processes		A	R	C	C	
Define roles and responsibilities	A	R	C	I	I	
Establish policies, procedures and standards for data quality	A	R	R	C	C	
Create a business data dictionary		A	C	C	R	
Define information systems support		I	A	C	R	
...						

R – Responsible; A – Accountable; C – Consulted; I – Informed

Table 24. RACI chart for data quality management tasks illustrates how these five actors should be included in the process of corporate data quality management (Wende, 2007, p. 420).

This subsection explored the theory of administration of the data quality management. The model clarifies the organizational needs in data quality management. The model gives an overview of the tasks that allow an organization to involve business aspect to the data quality management.

4.4.9 Corporate Data Quality Management Results

This subsection concludes the theory explored in section 4.4. The subsection sought to explore the research question: How should a corporation manage its data quality? The answer to this research question was sought through the means of systematic literature review that contrary to narrative literature review, applies a systematic and repeatable process to gather the material thus eliminating possible personal bias and lack of thoroughness. This approach adds value to the study because the studies included in this research always apply narrative literature review. The systematic approach narrows the existing research gap caused by the narrative literature reviews.

The section defined key concepts *corporate data philosophy, corporate data policy,*

corporate data strategy, data governance, data quality management, data owner, data quality methodology, data quality techniques, data quality tools, master data, master data quality management and *data steward*. The corporation needs to understand in order to manage the data assets.

After this the functions of corporate data quality management were explored through the model of Otto et al. (2012). This model defines six function groups that are divided into more specific functions. The function groups are: *Data Lifecycle Management, Metadata Management and Masterdata Modelling, Data Quality Assurance, Data Integration, Cross Functions* and *Administration.* This model forms a skeleton to which further theory has been added in order to explore the needed functions to manage the quality of data assets.

5. Discussion

This section discusses how a company should organize its data quality management so that it supports the business functions and creates value for the company. The research problem is studied through the defined three research questions by discussing every research question in their own separate subsections. The order of the subsections is the order of the research questions. After that the theoretical contribution of the study is discussed and after that the limitations and validity of this study are discussed. The research questions are:

1. How can a corporation define and assess its data quality?
2. What are the common problems in data quality management?
3. How should a corporation manage its data quality?

5.1 Corporate Data Quality Assessment

This section discusses the first research question: How can a corporation define and assess its data quality? The discussion is based on the material collected in the systematic literature review that can be seen in section 3. The results are explored in section 4.

Before a company can define its desired level of data quality, the relevant employees have to understand the concepts defined in the beginning of the literature review. Miller et al. (2001) and Lucas (2010) describe data, information and knowledge as interrelated concepts which form a chain that help the employees to understand the function of data in the organization and how it is the foundation of knowledge. Miller et al. (2001) emphasize the chained relationship of *data, information* and *knowledge* and how *information* and *knowledge* are more highly processed products made from *data.* There is a notion that if the *data* has poor quality, the *information* and *knowledge* are going to have the same problem due to the interrelated relationship.

The models of Wang (1998) and Parssian et al. (2004) help the relevant employees to understand data and information as products that have a process of creation, usage and deletion that can be compared to an industrial manufacturing process. The understanding of the product concept helps the employees to understand the processes that relate to data and to develop them in order to gain better data quality level. The model of information producers and consumers

helps the relevant employees to understand their roles with the data (Wang, 1998).

Otto et al. (2011) discuss about the lack of universal definition for data quality and conclude that "there is consensus in the specification of the term data quality as the data's fitness for use" (Otto et al. 2011, p.398). Also, Pipino et al. (2002) emphasize that the contextual nature of the data quality and note there is no "one size fits all" solution for defining data quality. After this, both Capiello et al. (2004) and Ofner et al. (2012) note that the data is a product of a process.

The notion of data being a contextual product of an information product manufacturing process is what ties together the themes of the listed authors in sections 4.2.1, 4.2.2 and 4.2.3. Capiello et al. (2004) emphasize the importance of understanding the manufacturing process behind the data because the manufacturing process creates the data and is ultimately the cause of the quality of the data. It can be concluded that the definition and assessment process can be started by studying the process that creates the data.

After understanding the process of data creation, the company needs to look for a methodology to further define and assess it. Batini et al. (2009) make the notion that there are many data quality methodologies. They have summarized the methodologies in Table 11. Their model can help companies to compare the existing data quality methodologies by comparing the different processes included in the methods. This comparison can help companies to acquire and implement the most suitable data quality methodology.

Finally the model of Wang (1998) was presented as an example methodology that can be used in defining and assessing the data quality. The work of Wang and Strong was used because it is widely referred to in the research material that was gathered in the systematic literature review. It defines and assesses data quality through four identified quality categories (*Intrinsic, Accessibility, Contextual* and *Representational*) and their dimensions. After this, a circular model of data quality management is presented.

A company's process of defining and assessing its data quality can be started by understanding the relationship between *data, information* and *knowledge* and understanding how important the quality of *data* is. After this, the company needs to understand that the data

is a product of an information product manufacturing process that is comparable to an industrial manufacturing process. The process affects the data and its quality. After understanding the background of the data, the company can identify its own information needs and compare different data quality methodologies and choose one that fits its needs optimally.

This study presented a widely used model developed by Wang (1998). The method enables a corporation to define its data quality through four data quality categories (*Intrinsic, Accessibility, Contextual* and *Representational*) and their dimensions. This method describes the process of data quality assessment as a continuous and iterative process. Even though this model is widely used, the company should always keep in mind the point emphasized by Otto et al. (2011), Pipino et al. (2002), Capiello et al. (2004) and Batini et al. (2009) that data is a product of a manufacturing process and that the definition and assessment highly depends on the process and the environment around the data and the information needs of the company.

5.2 Common Problems in Data Quality Management

This section discusses the second research question: What are the current problems in data quality management? The discussion is based on the material collected in the systematic literature review that can be seen in section 3. The results are explored in section 4.

The two main problems identified by the authors in section 4.3.1 are the lack of preventive data quality management and the lack of business involvement in the data quality management. The material emphasizes that there is a need to do a shift from reactive data quality management to preventive data quality management by changing the processes and data quality surveillance. There is also a clear emphasis on the need to include business aspect to the data quality management processes.

Otto et al. (2007), Haug et al. (2013), Silvola et al. (2011), Lucas (2010) and Hüner et al. (2011) explore various problems in current organizational corporate data quality management process. Subsection 4.2.1 presents a list of the most common problems. These listings can be used as checklists in exploring the data quality problems in an organization. This study emphasizes that the listed problems can eventually be linked to the two major problems: the lack of business involvement and the need for a shift from reactive to preventive

data quality management.

5.3 Managing the Corporate Data Quality

This section discusses the third research question: How should a corporation manage its data quality? The discussion is based on the material collected in the systematic literature review that can be seen in section 3. The results are explored in section 4.

Before the research question can be discussed, it is important to link the research question to the previous two research questions. Before a corporation can start discussing about the management of its data quality, it has to understand the issue of defining and assessing the data quality and understand the current problems of the data quality. As discussed in section 5.1, the corporation needs to understand its processes that create the data and understand how to define their desired level of data quality. After this the problems with the data quality need to be identified.

The section 5.2 identified two major problems in the modern data quality management: the lack of business involvement and the need to make the shift from reactive to preventive data quality management. Otto et al. (2007) notes the lack of coherent strategy in order to solve the current data quality issues. To this, Lucas (2010) emphasizes that the corporations need to create a strategy that involves the entire organization to the data quality management. Lucas (2010) defines important terms of *corporate data philosophy, corporate data policy, corporate data strategy, data governance, data quality management, data owner, data quality methodology, data quality techniques, data quality tools, master data, master data quality management* and *data steward*. These concepts allow a company to define an entire organizational change that allows the creation of the principles and guidelines that are needed to drive the data quality management to the desired direction. Defining these concepts can be seen as the first phase in managing the data quality.

The next step in managing the corporate data quality is to define functions to structure the management of the data quality management. This research uses the model of six function groups developed by Otto et al. (2012) because it addresses the identified problems in data quality management. The model involves employees organization-wide which involves the

business aspect to the data quality management thus creating a more strategic way to manage the data quality. The model also promotes more preventive measures in data quality management.

All of the six function groups can be seen to address the lack of business involvement. According to Ofner et al. (2012) the data lifecycle management improves the coordination of the data which in turn supports the business involvement. Data lifecycle management also supports the development of the status of the data in the organization and makes it more of a business asset (Ofner et al., 2012). Linking the data quality requirements to business process management also supports the linkage of data quality to business needs (Ofner et al., 2012). The training of the employees also helps them to involve the business requirement (Huang, 2015).

Metadata management aides the business with the business metadata repository that would be constructed with the help of business and technical employees (Hüner et al., 2011). The data quality assurance function may involve business needs to the data quality management by involving the business needs into the data quality evaluation techniques (Derby et al., 2011). Elgammal et al. (2012) also note that data quality assurance function can link the business requirements to data quality management through controls.

Cheng and Chang (1998) note that the data integration through data warehousing links the business needs and data quality by providing more quality data to business-related end users. Haug et al. (2009) support this by defining data quality dimensions in the ERP data. The model of Xu et al. (2002) may help the company to involve business aspect in the implementation phase of the ERP system. The functions in the cross functions function group help companies to monitor and control the business involvement in the data quality management (Otto et al., 2012). Administration function group brings clarity to the management of data quality project. It clarifies the roles of the stewards and involves a business data steward who involves the business aspect into the data quality management process (Wende, 2007).

Data lifecycle management drives the data quality management to be more preventive than reactive by coordinating the lifecycle better instead of reactively fixing the data (Ofner et

al., 2012). Also, the model of bringing data quality aspect into the business process management helps to prevent poor quality data from being entered into the information systems (Ofner et al., 2012). Training of the relevant employees also prevents the poor quality data from being entered into the system (Huang, 2015).

The purpose of metadata management is to prevent poor quality metadata and ensure standards for it (Otto et al., 2012). Aljumaili, et al (2016) introduce possible set of constraints for metadata that prevent poor quality metadata. The data quality controls in the data quality assurance function group can also prevent poor quality data (Elgammal et al., 2012).

Different data quality assessment techniques with the data integration prevent the poor quality data from being entered into the system (Otto et al., 2012; Haug et al., 2009). The functions in the cross functions function group help companies to control the preventive processes (Otto et al., 2012). Organizing accountabilities for data quality management prevents poor quality data being entered to the system by clearly defining the roles and responsibilities in the process of data quality management (Wende, 2007).

In order to manage its data quality, a corporation needs to form their own *corporate data philosophy, corporate data policy, corporate data strategy, data governance* in order to coherently involve the entire organization and make sure that the business aspect is involved. After this the corporation needs to decide the functions that are going to be deployed in the organization to manage the data. The six function groups developed by Otto et al. (2012) can be used as a checklist of the different aspects that need to be considered in the process.

5.4 Summary of the Results

This study reviewed the existing studies about how a company should organize its data quality management so that it supports the business functions and creates value for the company. This topic was studies through the three research questions that were discussed in sections 5.1, 5.2 and 5.3. Figure 10 summarizes the key findings and topics of the systematic literature review. It can be used as a quick check list in order to review the important issues in corporate data quality management.

The key takeaway in data quality assessment is to understand that the data is a product

made in a process and that data quality is data's "fitness for use". The most common problems in data quality management are the lack of business aspect and the lack of preventive measures in the management of the data quality. An organization can manage its data quality by ensuring that the defined six functions are being conducted.

EFFECTIVE CORPORATE DATA QUALITY MANAGEMENT

Data Quality and Assessment

Gain understanding of the needs of the data consumers
Gain understanding of the processes behind the data
Data quality definition --> data's "*fitness for use* "
Data quality methodology assessment table Batini et al. (2009)

Common Problems in Data Quality Management

Understand two main issues in modern data quality management
1. Lack of business aspect in the data
    Ignored needs of the data consumers
2. Lack of preventive data quality measures
    Issues are fixed after they have occured

Corporate Data Quality Management

Design and implement processes to address the definition of data quality and the defined data quality problems according to the six function groups:
1. Master Data Lifecycle Management
    Data Creation, Data Maintenance, Data Deactivation, Data Archiving
2. Metadata Management and Master Data Modeling
    Data Modeling, Model Analysis, Metadata Management
3. Data Quality Assurance
    Data Analysis, Data Enrichment, Data Cleansing
4. Master Data Integration
    Data Import, Data Transformation, Data Export
5. Cross Functions
    Automation, Reports, Search, Workflow Management
6. Administration
    Data History Management, User Management

Figure 10. The key findings and topics of the systematic literature review.

5.5 Theoretical Contributions

This study conducted a systematic literature review that analyzed the gathered material. This section discusses the theoretical contributions of the study. First, the analysis of the research

material is discussed and then the research questions are discussed.

Cooper et al. (2004) note that the systematic analysis of the gathered research material is added value. The analysis of the research material is presented in section 4.1. Figure 3 illustrates that the amount of studies has grown between 1990 and 2016 indicating that the issue of data quality is considered more important issue that needs to be addressed. Table 4 indicates that data quality studies are being published by various sources that also indicates the importance of the topic.

This study studied the methods that were used in the included material. It is essential to notice that the issue of data quality and its management is studied mainly through empirical case studies and narrative literature reviews. These methods do not explore the existing material with a replicable process, but there is a possibility that the chosen material in the studies are affected by personal bias and lack of thoroughness. This indicates that there is a need for a systematic literature review that reviews the material in a systematic, replicable process. Section 4.1 can be considered to be a contribution to the data quality literature because it gives an overview of trends in a sample of included studies.

Another important point of the literature is described in Table 5 that explores the theoretical background behind the articles. The theory of Total Data Quality Management of Wang (1998) is the most influential theoretical background in the sample. It must also be noted that it is also referred to by almost every study in the sample. The study of Batini et al. (2009) compares the different methodologies and illustrates that the methodology of Wang (1998) is covers various steps in data quality assessment. This comparison may clarify the popularity of Wang (1998).

There are also some problems in the high influence of the theory of Total Data Quality Management by Wang (1998). It is notable that no such studies that would have challenged the theory of Total Data Quality Management, were found during the literature inclusion. This study notes that there may be a risk in the situation in which the entire science community keeps praising the almost two decades old theory without developing it further or questioning some of the aspects of the study. The possible shortcomings of the theory of Total Data Quality Management (Wang, 1998) may get transferred to the further research if the points of Wang

(1998) are not questioned.

As a further research topic, this study suggests further studies on the theory of Total Data Quality Management. This includes more case studies that would test the practical aspects and functionality of the theory. There could also be possible further research that would study the theory developed in 1998 in the environment of 2017 thus studying the effect of the updates in technology.

Section 4.2 answers the research question: How should a corporation define and assess its data quality? Lucas (2010) notes that the current literature lacks clarity and should be reviewed to define and assess the data quality. The theoretical contribution of this section is that it gives an overview of the different theories that are linked to the research question through the systematic literature review. It also combines various data quality tools and presents how they can be used to define and assess the data quality of a corporation.

Section 4.3 answers the research question: What are the current problems in data quality management? Otto et al. (2007) discuss that the current literature on data quality does not answer the current challenges of the data quality management. The theoretical contribution of the section is that it wraps up the literature on the current problems and presents a short list of them. The key theoretical issue in the section is the emphasis that the data quality is still considered as a highly technical issue and that the scientific community still struggles at creating the shift from technical to a business issue.

Section 4.4 answers the research question: How should a corporation manage its data quality? Lucas notes that there is a need to gather the existing literature to clearly answer the question. Otto et al. (2007) note that the data quality management needs to answer the identified issues. The section combines various existing theories by linking them to the model of Otto et al. (2012). The theoretical contribution of the combined theories is that it forms a detailed and developed set of functions that can be deployed to manage the data quality and thus develops the model of Otto et al. (2012) further into details.

The material used to answer the three research questions can be seen as a theoretical guide in the management of the corporate data quality. The material used in the first research question gives an overview of a set of tools that can be used to define and assess the data quality of a

corporation. The material related to the second research question can be seen as a theoretical checklist that can be used to identify root causes of the data quality issues in a corporation. The material related to the third research question gives a theoretical overview of the tasks related to corporate data quality management. The theories combined have been derived from a process of systematic literature review. This is summarized in section 5.4 that gives a managerial overview of the theoretical contribution of the study in Figure 10.

5.6 Implications to Management

The literature about the data quality is full of estimates about the expenses of the poor data quality which indicate the importance of data quality studies. There is a need for various data quality studies to ease the corporate work. This section discusses about the key managerial implications of this research.

This study can be used as a theoretical manual in practical corporate data quality management. The study has gathered the existing scientific material in a replicable systematic literature review and reviewed it to answer the three inter-related research questions that allow a corporation to review the current state of its data quality and improve it. This study can give the employees responsible of corporate data quality management an overview of how to define and assess the data quality, a description of common problems in data quality management and a set of functions to develop the data quality.

Section 4.2. gives the data quality employees a collection of tools to define and assess the data quality of the corporation. The section gives an overview of the data being a product of a process and it helps the data quality employees to understand the process behind the data. The data quality employees can assess the possible data quality methodologies with the model of Batini et al. (2009) and define the methodology most optimal for their use.

Section 4.3 gives an overview of the most common problems in the data quality management. The identified two main problems of data quality management: the lack of business aspect and the lack of preventive measures can be considered to be the root cause of many data quality issues in a corporation. The problems should be actively searched and recognized inside the organization.

Section 4.4 gives corporations a set of functions that can be used to control the quality

of the data assets. The section can be used as a checklist of functions that need to be deployed inside the organization in order to manage the data quality. The theory gathered in the section can also be used in determining different tasks in the data quality management.

5.7 Discussion on the Limitations, Reliability and Validity

This section discusses the limitations, reliability and validity of this study. The goal of the study was to conduct a replicable systematic literature review that would explore the existing literature on data quality management and answer the defined research questions. Even though the research followed the principles of the systematic literature review, it must be recognized that there are no perfect research methods and that the study has limitations.

The process of the systematic literature review itself contains important limitations that need to be discussed. In the beginning of the study, a scoping study was conducted. The purpose of the scoping study was to give an overview of the themes and trends in the literature. The scoping study can also be seen as a limitation to the study, because the themes and trends indicated during this phase have a large effect on the later phases of the study. For example, the search terms are defined based on the material discovered in the scoping study.

Considering the scoping study does not cover the entire literature in a systematic way, there is a possibility that some aspects in the studied issue get over-emphasized due to the choice of search terms. This is a threat to the systematic process that is one of the aims of the systematic literature review. This study recognizes the issue and has mitigated the risk of poor search term definition by defining seven different search terms that explore different aspects of the existing literature better that one or two search terms. Another way this study mitigated the risk was to use three different studies in the scoping phase. This ensured that the scoping study was not effected by only one study that could have turned the focus of the research into a wrong way.

The database and the search algorithms form a notable limitation to the systematic process of the systematic literature review. The search algorithms may emphasize some studies too much compared to others. This study mitigated the risk by employing four different databases so that the search process would not be dependent on only one set of algorithms.

Another important limitation of the study was the large amount of research material

that needed to be reviewed in limited resources as can be seen in subsection 3.1.2. The material needed to be limited due to the extensive first stage results. This study solved the issue by including the first 500 search results in each used database. If the principles of systematic literature review are considered, this approach does not theoretically cover the entire literature in the most systematic way.

This study recognizes this theoretical problem. This issue has been mitigated with the relevancy function of the databases. The databases sort the material on the basis of its relevance to the applied search terms. Another measure that was used in order to mitigate the risk was the empirical notion made about the material that the first 500 search results usually covered the most relevant material.

This study recognizes that the usage of the Snowball method can be seen as a risk to the systematic process and to the aim of eliminating the personal bias because it allows linking the material to further material on the basis of the judgement of the researcher without including the material in a formal, defined inclusion process. This theoretically allows the researcher to use his or her own personal bias in including the material. Also, the method can be seen as a threat to the aimed thoroughness of the process.

This study recognizes the risk related to the usage of the Snowball method, but also notes that the method is needed in order to cover material that has been possibly excluded in the systematic literature review due to the restriction of the first 500 search results. Also, the search results of the systematic literature review did not cover some of the aspects of the studied topic, for example compliance. Finally, the risk of personal bias is mitigated with the small amount of Snowball-linked material. Only four studies were included to the pool of 31 studies discovered in the systematic literature review.

The material explored in the study contain some issues that may pose a limitation to the reliability and validity of the study. The research methods chosen be Wang and Strong (1996) and Haug et al. (2013) Wang and Strong (1996) have conducted their study as a survey research. Survey research method is widely used and accepted research method, but it may have some troublesome elements that have to be discussed in this section. The first survey of Wang and Strong (1996) has been sent to 25 industry workers and 112 MBA students and the

second survey was sent to 1500 MBA program alumni. Wang and Strong (1996) chose to approach an MBA alumni because they considered them to have a stable proportion of various industries involved. Wang and Strong do not explain the choice of their chosen sample sizes nor the proportions of different industries involved in the research. Slater and Atauhene-Gima (2015) emphasize the importance of correct sample size selection in a survey research and the fact that Wang and Strong (1996) do not explain their choice, must raise skepticism over the research.

Haug et al. (2013) have carried out their research by surveying the financial managers of the chosen survey companies because they are considered to have a cross-organizational and neutral view on the entire organization. Slater and Atauhene-Gima (2015) emphasize that the respondent of the survey has to be as knowledgeable as possible. The idea of financial managers having a cross-organizational view on their organizations can be questioned because it is based on the assumption and generic description of the task. The organizations have not been studied individually to determine the best informant.

Section 4.4 offers one theoretical solution to the data quality management of a company. It integrates further theory to a set of six highly inter-related function groups and functions as a checklist of the possible aspects that should be taken into consideration when managing the corporate data quality. It should be emphasized that considering the inter-relatedness of the function groups, other models can be developed by considering other categorization possibilities.

Another important notion of the usability of the theoretical model is that the model does not take into account the possible differences posed by specific business requirements of different fields of business. There are various empirical studies on the topic that have been conducted in various types of companies and educational, medical, military and research institutions. The academics often quote each other without emphasizing the differences in the fields of business that may pose some differences in requirements to data quality. The requirements for data quality posed by specific business fields (e.g. banking, telecommunications, retail) can be considered to be a further research topic.

6. Conclusion

This thesis reviewed the scientific literature about corporate data quality management. This section gives the final conclusion of the researched topic. Also, the managerial implications are discussed in section 6.1.

The focus of this study was in defining and assessing the data quality, discuss the current problems related to it and review models and case studies that have been written on the issue. The data quality was defined through "fitness for use" concept that has been established in the academic discussion on the topic. (Otto et al., 2012) The current problems of corporate data quality management consist of the lack of business involvement and need for a shift from reactive to preventive corporate data quality management. At the end a model for corporate data quality management was constructed by combining the relevant models and case studies.

This thesis works as a review on the scientific discussion on corporate data quality. It assesses a set of tools to assess data quality through data quality attributes. It reviews and indicates the current major problems in the field of corporate data quality. After this, it reviews organization-wide theories that explores the issue through the model of Otto et al. (2012) and combines various articles and case studies into a single model of six distinct function groups. This set of theories can be used in planning the organization-wide data quality strategy.

Overall, it can be said that the issue of corporate data quality is an important problem point in various modern companies and the research on the topic is crucial for the business success of these companies. The research on the topic should be continued because of the clear need for clarity in corporate data quality models and the explored weaknesses in them. Better data quality management can also be considered as an important tool in managing more and more data-driven society.

6.1 Further Research Topics

Various studies have proved that data quality issues are very common and cost money for companies (Lucas, 2010). The listings of the common data quality problems could be studied further by developing case studies around them in real organizations. The prospective case study companies could be approached, the problem could be identified and eventually, a

solution model for the problem could be developed.

An important feature to add to the model in the further research are empirical studies. It must be emphasized that the model integrates so many aspects and functions that studying all of them in one empirical study would be problematic. The different aspects and function groups of corporate data quality management can be studied through interviewing relevant specialists, surveying samples of information producers and consumers, and creating case studies around the functions.

Considering the amount of companies and other organizations that struggle with the issues with data quality management, future thesis writers could contact different companies and offer a thesis project on data quality for them. The topics and concentration points for these empirical case studies could be the possible requirements of the specific business types. The theoretical model presented in Figure 10 could be used as the basis for the literature review and as a tool to define appropriate hypotheses.

Another further important research point for the academic community around data quality management would be to define some of the theoretical concepts better. Currently various concepts and definitions are used highly interchangeably. This would ease the work of various people who currently work on these issues because currently the interchangeability causes misinterpretations among the researchers.

The systematic literature review also revealed that the work of Wang (1998) has had a major influence in the data quality studies. The further research on the work of Wang (1998) should be conducted in order to discover possible issues in the data quality management studies. The theory should be questioned and further case studies could be made on the basis of it.

References

Aljumaili, Mustafa, Ramin Karim, and Phillip Tretten. "Metadata-Based Data Quality Assessment." *VINE Journal of Information and Knowledge Management Systems* 46.2 (2016): 1-27.

Atkinson, Rowland, and John Flint. "Accessing Hidden and Hard-to-Reach Populations: Snowball Research Strategies." *Social Research Update* Summer 2001 33 (2001): 1-8.

Baltar, Fabiola, and Ignasi Brunet. "Social Research 2.0: Virtual Snowball Sampling Methods Using Facebook." *Internet Research* 22.1 (2012): 57-74.

Bartunek, Jean M., Philip Bobko, and N. Venkatraman. "Towards Innovation and Diversity in Management Research Methods." The Academy of Management Journal 36.6 (1993): 1362-373.

Batini, Carlo, Cinzia Cappiello, Chiara Francalanci, and Andrea Maurino. "Methodologies for Data Quality Assessment and Improvement." ACM Computing Surveys 16th ser. 41.3 (2009): 1-50.

Cappiello, Cinzia, Chiara Francalanci, and Barbara Pernici. "Data Quality Assessment from the User's Perspective." IQIS Proceedings (2004): 68-73.

Cheng, Paul S., and Pintsang Chang. "Transforming Corporate Information into Value through Data Warehousing and Data Mining." *Aslib Proceedings* 50.5 (1998): 109-13.

Cooper, B. S., S. P. Stone, C. C. Kibbler, B. D. Cookson, J. A. Roberts, G. F. Medley, G. Duckworth, R. Lai, and S. Ebrahim. "Isolation Measures in the Hospital Management of Methicillin Resistant Staphylococcus Aureus (MRSA): Systematic Review of the Literature." BMJ 329 (2004): 1-8.

Croom, Simon, Pietro Romano, and Mihalis Giannakis. "Supply chain management: an analytical framework for critical literature review." European Journal of Purchasing & Supply Management 6 (2000): 67-83.

Deloitte Touche Tohmatsu. *Telecommunications Predictions TMT Trends 2009*. London: Creative Studios Deloitte, 2009. Print.

Derby, Dustin C., Andrea Haan, and Kurt Wood. "Data Quality Assurance: An Analysis of Patient Non-response." *International Journal of Health Care Quality Assurance* 24.3 (2011): 198-210.

Elbireer, Ali, Julie Le Chasseur, and Brooks Jackson. "Improving Laboratory Data Entry Quality Using Six Sigma." *International Journal of Health Care Quality Assurance* 26.6 (2013): 496-509.

Elgammal, Amal, Oktay Turetken, and Willem-Jan Van Den Heuvel. "Using Patterns for the Analysis and Resolution of Compliance Violations." *International Journal of Cooperative Information Systems* 21.1 (2012): 31-54.

Falge, Clarissa, Boris Otto, and Hubert Österle. "Data Quality Requirements for Collaborative Business Processes." *2012 45th Hawaii International Conference of System Sciences* (2012): 4316-325.

Gammelgaard, Britta. "Schools in logistics research? A methodological framework for analysis of the discipline ." International Journal of Physical Distribution & Logistics Management 6th ser. 34 (2004): 479-91.

Glowalla, Paul, and Ali Sunyaev. "ERP System Fit - an Explorative Task and Data Quality Perspective."*Journal of Enterprise Information Management* 27.5 (2014): 668-86.

Greenhalgh, T., and R. Taylor. "Papers That Go beyond Numbers (qualitative Research)." *BMJ Clinical Research* 315.7110 (1997): 740-43

Haug, Anders, Jan Stentoft Arlbjorn, Frederik Zachariassen, and Jakob Schlichter. "Master Data Quality Barriers: An Empirical Investigation." *Industrial Management & Data Systems* 113.2 (2013): 234-49.

Haug, Anders, Arlbjorn, Jan Stentoft, and Pedersen, Anne. "A Classification Model of ERP System Data Quality." *Industrial Management & Data Systems* 109.8 (2009): 1053-068.

Huang, Hong. "Domain Knowledge and Data Quality Perceptions in Genome Curation Work." *Journal of Documentation* 71.1 (2015): 116-42.

Hurley, Margaret A., and Rod Harris. "Facilitating Corporate Knowledge: Building the Data Warehouse." *Information Management & Computer Security* 5/5 (1997): 170-74.

Hüner, Kai M., Boris Otto, and Hubert Österle. "Collaborative Management of Business Metadata." *International Journal of Information Management* 31 (2011): 366-73.

Hüner, Kai M., Martin Ofner, and Boris Otto. "Towards a Maturity Model for Corporate Data Quality Management." SAC 09 Proceedings of the 2009 ACM Symposium on Applied Computing (2009): 231-38.

John, S. A. "Data Integration in a GIS - the Question of Data Quality." *Aslib Proceedings* 45.4 (1993): 109-19.

Kwon, Ohbyung, Namyeon Lee, and Bongsik Shin. "Data Quality Management, Data Usage Experience and Acquisition Intention of Big Data Analytics." *International Journal of Information Management* 34 (2014): 387-94.

Lillrank, Paul. "The Quality of Information." *International Journal of Quality & Reliability Management* 20.6 (2003): 691-703.

Lucas, Ana. "Towards Corporate Data Quality Management." *Portuguese Journal of Management Studies* XV.2 (2010): 173-95.

Ma, Catherine, David C. Chou, and David C. Yen. "Data Warehousing, Technoloy Assessment and Management." *Industrial Management & Data Systems* 100/3 (2000): 125-34.

Miller, Bob, Mary Ann Malloy, Ed Masek, and Chris Wild. "Towards a Framework for Managing the Information Environment." *Information Knowledge Systems Management* 2 (2001): 359-84.

Morgan, Gareth, and Linda Smircich. "The Case of Qualitative Research." The Academy of Management Review 5.4 (1980): 491-500.

Ofner, Martin H., Boris Otto, and Hubert Österle. "Integrating a Data Quality Perspective into Business Process Management." *Business Process Management Journal* 18.6 (2012): 1036-1067.

Ofner, Martin Hubert, Kevin Straub, Boris Otto, and Hubert Oesterle. "Management of the Master Data Lifecycle: A Framework for Analysis." Journal of Enterprise Information Management 26.4 (2013): 472-91.

Otto, Boris, Kai M. Hüner, and Hubert Österle. "Toward a Functional Reference Model for

Master Data Quality Management." Info Syst E-bus Manage 10 (2012): 395-425.

Otto, Boris, Kristin Wende, Alexander Schmidt, and Philipp Osl. "Towards a Framework for Corporate Data Quality Management." ACIS Proceedings Paper 109 (2007): 916-26.

Parssian, Amir, Sumit Sarkar, and Varghese S. Jacob. "Assessing Data Quality for Information Products: Impact of Selection, Projection, and Cartesian Product." Management Science 50.7 (2004): 967-82.

Pipino, Leo L., Yang W. Lee, and Richard Y. Wang. "Data Quality Assessment." Communications of the ACM 45.4ve (2002): 211-18.

Qu, Sandy Q., and John Dumay. "The Qualitative Research Interview." *Qualitative Research in Accounting and Management* 8.3 (2011): 238-64.

Scarisbrick-Hauser, AnneMarie, and Christina Rouse. "The Whole Truth and Nothing but the Truth? The Role of Data Quality Today." *Direct Marketing: An International Journal* 1.3 (2007): 161-71.

Silvola, Risto, Olli Jaaskelainen, Hanna Kropsu-Vehkapera, and Harri Haapasalo. "Managing One Master Data - Challenges and Preconditions." *Industrial Management & Data Systems* 111.1 (2011): 146-62.

Slater, Stanley F., and Kwaku Atuahene-Gima. "Conducting Survey Research in Strategic Management." *Published Online* (2015): 227-249. *Emerald Insight*. Web. 20 July 2016. <http://www.emeraldinsight.com/doi/full/10.1016/S1479-8387%2804%2901108-7>.

Strong, Diane M., Yang W. Lee, and Richard Y. Wang. "10 Potholes in the Road to Information Quality." *Cybersquare* August (1997): 38-46.

Strong, Diane M., Yang W. Lee, and Richard Y. Wang. "Data Quality in Context." *Communications of the ACM* 40.5 (1997): 103-10.

Tranfield, David, David Denyer, and Palminder Smart. "Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review." British Journal of Management 14 (2003): 207-22.

Wand, Yair, and Richard Y. Wang. "Anchoring Data Quality Dimensions in Ontological Foundations." *Communications of the ACM* 39.11 (1996): 86-95.

Wang, Richard Y. "A Product Perspective on Total Data Quality Management." Communications of the ACM 41.2 (1998): 58-65.

Wang, Richard Y., and Diane M. Strong. "Beyond Accuracy: What Data Quality Means to Data Consumers." *Journal of Management Information Systems* 12.4 (1996): 5-34.

Wang, Richard Y., Veda C. Storey, and Christopher P. Firth. "A Framework for Analysis of Data Quality Research." IEEE Transactions on Knowledge and Data Engineering 7.4 (1995): 623-40.

Wende, Kristin. "A Model for Data Governance - Organising Accountabilities for Data Quality Management." ACIS 2007 Proceedings Paper 80 (2007): 417-25.

Xu, Hongjiang, Jeretta Horn Jord, Noel Brown, and G. Daryl Nord. "Data Quality Issues in Implementing an ERP." *Industrial Management & Data Systems* 102.1 (2002): 47-58.

Appendix

The theories used in the gathered studies. The theory is indicated in the Theory Used column.

Year	Authors	Title	Journal	Approach	Focus Area	Theory Used
1993	John, S.A.	Data integration in GIS - the question of data quality	Aslib Proceedings	Empirical study: case study	Data Integration	Obtaining Information on Quality Digital Data (1986)
1995	Wang, Richard Y.;Storey, Veda C.and Firth, Christopher P.	A Framework for Analysis of Data Quality Research Data Quality Research	IEEE Transactions on Knowledge and Data Engineering	Narrative literature review	Data Quality	What Data Quality Means to Data Consumers (Wang, Strong, Guarascio, 1994)
1996	Wand, Yair and Wang, Richard Y.	Anchoring Data Quality Dimensions in Ontological Foundations	Communications of the ACM	Narrative literature review	Data Quality	A Framework for Analysis of Data Quality Research Data Quality Research (Wang, Stroey and Firth, 1995)
1996	Wang, Richard Y.; Strong, Diane	Beyond Accuracy: What Data Quality Means to Data Consumers	Journal of Management Information Systems	Empirical study: survey research	Data Quality	Anchoring Data Quality Dimensions in Ontological Foundations (Wand and Wang, 1996)
1997	Hurley, Margaret A. and Harris, Rod	Facilitating corporate knowledge: building the data warehouse	Information Management & Computer Security	Narrative literature review	Data Warehouse	N/A
1998	Cheng, Paul S. and Chang, Pintsang	Transforming corporate information into value through data warehousing and data mining	Aslib Proceedings	Empirical study: case study	Data Warehouse	N/A
1998	Wang, Richard Y	A Product Perspective on Total Data Quality Management	Communications of the ACM	Empirical study: survey research	Data Quality	A Framework for Analysis of Data Quality Research Data Quality Research (Wang, Stroey and Firth, 1995)
2000	Ma, Catherine; Chou, David C. ; Yen, David C.	Data warehousing, technology and assessment	Industrial Management & Data Systems	Narrative literature review	Data Warehouse	The entity-relationship model - towards a unified view of data (Chen, 1976)

2001	Miller, Bob; Malloy, Mary Ann; Masek, Ed and Wild, Chris	Towards a Framework for Managing the Information Environment	Information Knowledge Systems Management	Empirical study: case study	Information Management	Mastering the information and knowledge environment (Davenport and Prusak, 1997)
2002	Pipino, Leo L; Lee, Yang W.; Wang, Richard	Data Quality Assessement	Comunications of the ACM	Empirical study: case study	Data Quality	Total Data Quality Management (Wang, 1998)
2002	Xu, Hongjiang; Nord, Jeretta Horn; Brown, Noel; Nord, G. Daryl	Data quality issues in implementing ERP	Industrial Management & Data Systems	Empirical study: case study	Database Management	Anchoring Data Quality Dimensions in Ontological Foundations (Wand and Wang, 1996)
2003	Lillrank, Paul	The Quality of Information	Interantionl Journal of Quality and Reliability Management	Narrative literature review	Data Quality	Improving Data Warehouse and Business Information Quality (English, 1999)
2004	Capiello, Cinzia; Capiello, Cinzia; Pernici, Barbara	Data quality assessment from the user's perspective	IQIS '04 Proceedings of the 2004 international workshop on Information quality in information systems	Narrative literature review	Data Quality	Total Data Quality Management (Wang, 1998)
2004	Parssian, Amir; Sarkar, Sumit ; Jacob, Varghese S.	Assessing Data Quality for Information Products: Impact of Selection, Projection, and Cartesian Product	Management Science	Empirical study: case study	Data Quality	Total Data Quality Management (Wang, 1998)
2007	Otto, Boris; Wende, Kristin; Schmidt, Alexander and Osl, Philipp	Towards a Framework for Corporate Data Quality Management	ACIS Proceedings	Narrative literature review	Data Quality	Total Data Quality Management (Wang, 1998)
2007	Scarisbrick-Hauser, AnneMarie, and Rouse, Christina	The Whole Truth and Nothing but the Truth? The Role of Data Quality Today	Direct Marketing: An International Journal	Narrative literature review	Data Quality	Improving Data Warehouse and Business Information Quality (English, 1999)
2007	Wende, Kristin	A Model for Data Governance - Organising Accountabilities for Data Quality Management	ACIS 2007 Proceedings	Narrative literature review	Corporate Data Quality Management	Total Data Quality Management (Wang, 1998)

2009	Batini, Carlo; Capiello, Cinzia; Ciara Francalanci, Ciara and Maurino, Andrea	Methodologies for Data Quality Assessment and Improvement	ACM Computing Surveys	Narrative literature review	Data Quality	Total Data Quality Management (Wang, 1998)
2009	Haug, Anders; Stentoft, Jan Arlbjorn; Pedersen, Anne	A Classification Model of ERP System Data Quality	Industrial Management & Data Systems	Empirical study: case study	ERP Data Quality	Anchoring Data Quality Dimensions in Ontological Foundations (Wand and Wang, 1996)
2009	Hüner, Kai M.; Martin Ofner, Martin and Otto, Boris	Towards a Maturity Model for Corporate Data Quality Management	SAC 09 Proceedings of the 2009 ACM Symposium on Applied Computing (2009)	Empirical study: case study	Corporate Data Quality Management	Total Data Quality Management (Wang, 1998)
2010	Lucas, Ana	Towards Corporate Data Quality Management	Portuguese Journal of Management Studies	Empirical study: case study	Corporate Data Quality Management	Improving Data Warehouse and Business Information Quality (English, 1999)
2011	Derby, Dustin C.; Andrea Haan, Andrea and Wood, Kurt	Data quality assurance: an analysis of patient non-response	International Journal of Health Care Quality Assurance	Empirical study: survey research	Data Quality Assurance	Medical
2011	Hüner, Kai M.; Otto, Boris and Österle, Hubert	Collaborative Management of Business Metadata	International Journal of Information Management	Empirical study: case study	Metadata Management	Business Metadata Management (Huner, 2011)
2011	Silvola, Risto; Jaaskelainen, Olli; Kropsu-Vehkapera, Hanna and Haapasalo, Harri	Managing one master data - challenges and preconditions	Industrial Management & Data Systems	Empirical study: in-depth interviews	Masterdata management	Methodologies for Data Quality Assessment and Improvement (Batini, Capiello, Francallacci, Maurino, 2009)
2012	Elgammal, Amal; Turetken, Oktay and Van Den Heuvel, Willem-Jan	Using Patterns for the Analysis and Resolution of Compliance Violations	International Journal of Cooperative Information Systems	Empirical study: case study	Regulatorly Compliance	Compliance
2012	Falge, Clarissa; Otto, Boris; Österle, Hubert	Data Quality Requirements of Collaborative Business Processes	2012 45th Hawaii International Conference on System Sciences	Empirical study: case study	Corporate Data Quality Management	Anchoring Data Quality Dimensions in Ontological Foundations (Wand and Wang, 1996)

2012	Ofner, Martin H; Otto, Boris; Österle, Hubert	Integrating a data quality perspective into business process management	Business Process Management Journal	Empirical study: case study	Process Management	Methodologies for Data Quality Assessment and Improvement (Batini, Capiello, Francallacci, Maurino, 2009)
2012	Otto, Boris; Huner. Kai M.; Österle, Hubert	Toward a functional reference model for master data quality management	Inf. Syst E-bus Manage	Empirical study: case study	Masterdata Management	Total Data Quality Management (Wang, 1998)
2013	Elbireer, Ali; Le Chasseur, Julie; Jackson, Brooks	Improving laboratory data entry quality using Six Sigma	International Journal of Health Care Quality Assurance	Empirical study: case study	Data Quality	Application of Six Sigma/CAP methodology: controlling blood-product utilization and costs (Neri, Mason, Demko, 2008)
2013	Haug, Anders; Albjorn, Jan Stentoft; Zachariassen, Frederik and Schlichter, Jakob	Master data quality barriers: an empirical investigation	Industrial Management & Data Systems	Empirical study: survey research	Masterdata management	Total Data Quality Management (Wang, 1998)
2013	Ofner, Martin H; Straub, Kevin; Otto, Boris and Österle, Hubert	Management of the masterdata lifecycle: a framework for analysis	Journal of Enterprise Information Management	Empirical study: case study	Masterdata management	Total Data Quality Management (Wang, 1998)
2014	Glowalla, Paul; Sunyaev, Paul	ERP System Fit - an Explorative Task and Data Quality Perspective	Journal of Enterprise Information Management	Empirical study: in-depth interviews	ERP Data Quality	Anchoring Data Quality Dimensions in Ontological Foundations (Wand and Wang, 1996)
2014	Kwon, Ohbyung; Namyeon Lee, Namyeon and Shin, Bongsik	Data Quality Management, Data Usage Experience and Acquisition Intention of Big Data Analytics	International Journal of Information Management	Empirical study: survey research	Masterdata Management	Business intelligence and analytics: Form big data to big impact (Chen, Chiang, Storey, 2012)
2015	Huang, Hong	Domain knowledge and data quality perceptions in genome curation work	Journal of Documentation	Empirical study: survey research	Data Quality	Beyond Accuracy: What Data Quality Means to Data Consumers (Wang and Strong, 1996)
2016	Aljumaili, Mustafa; Karim, Ramin and Tretten, Phillip	Metadata-based data quality assessment	VINE Journal of Information and Knowledge Systems	Empirical study: case study	Metadata, data quality	Data Quality in Context (Strong and Wang, 1997